Chapter 11

Rings and polynomials

Introduction

In Book A and Chapters 9 and 10, we studied properties of the set $\mathbb{Z}$ of integers. In this chapter and the next, we are going to take a step into the abstract to discover what 'factorisation' means for the elements of a set, equipped with a suitable definition for multiplication.

In this chapter, we will first explore the abstract definitions of a *ring* and a *field*, before studying *polynomials* where the coefficients of the polynomials are elements from a given field. It turns out that these 'polynomial rings' have many properties in common with the integers; for example we can define a division algorithm, and this enables us to develop the analogue of the 'highest common factor' for two polynomials. In the final section of this chapter, we will look more closely at polynomials with integer coefficients, and prove several results relating to the factorisation of a more general collection of polynomials, namely those with *rational* coefficients.

Chapter guide

This chapter has four sections, with each section relying, to an extent, on the content of the previous sections. The first section contains relatively few proofs, but is quite long and defines algebraic concepts such as rings and fields, which are required by the rest of this chapter as well as Chapter 12 and much of Book E. The second section is shorter than the other three, and defines concepts needed for the final two sections, although these themes also reoccur in later chapters. The last two sections are not built upon as much by later chapters as the first two, although the concepts they contain do reoccur, particularly in Chapter 12.

1 Rings

In Book B we learnt about groups, which are defined as a set of elements equipped with some operation $\circ$, and satisfying certain axioms given in Definition 1.1 of Book B, Chapter 5. When we think of the set of integers, however, there are two basic operations that we can use: addition, $+$, and multiplication, $\cdot$. We learnt in Chapter 5 that $(\mathbb{Z}, +)$ is an abelian group, but we can quickly verify that $(\mathbb{Z}, \cdot)$ is not: no integers apart from 1 and -1 have multiplicative inverses and so axiom G3 of Definition 1.1 does not hold.

What about subtraction and division? In a sense, we can think of these as the inverses to addition and multiplication.

We may ask which of the group axioms *do* hold for $(\mathbb{Z}, \cdot)$. If $m, n \in \mathbb{Z}$, then $m \cdot n \in \mathbb{Z}$ and so we have closure (axiom G1). Next, since $1 \cdot m = m \cdot 1 = m$, the element $1 \in \mathbb{Z}$ acts as the identity to confirm axiom G2. Finally, associativity of multiplication is a basic property of the integers and so G4 holds. The integers form a model for our definition of a ring.

The old German word *Ring* can mean 'association'; hence the terms 'ring' and 'group' have similar origins.

Definition 1.1 *Ring axioms*

Let R be a set and let $+$ and $\cdot$ be binary operations defined on R. Then $(R, +, \cdot)$ is a **ring** if the following axioms hold.

Axioms for addition:

R1 Closure For all $a, b \in R$,

$$a + b \in R.$$

R2 Associativity For all $a, b, c \in R$,

$$a + (b + c) = (a + b) + c.$$

R3 Additive identity There exists an additive identity $0 \in R$ such that, for all $a \in R$,

$$a + 0 = a = 0 + a.$$

R4 Additive inverses For each $a \in R$, there exists an additive inverse $-a \in R$ such that

$$a + (-a) = 0 = (-a) + a.$$

R5 Commutativity For all $a, b \in R$,

$$a + b = b + a.$$

Axioms for multiplication:

R6 Closure For all $a, b \in R$,

$$a \cdot b \in R.$$

R7 Associativity For all $a, b, c \in R$,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

R8 Multiplicative identity There exists a multiplicative identity $1 \in R$ such that, for all $a \in R$,

$$a \cdot 1 = a = 1 \cdot a.$$

Axioms combining addition and multiplication:

R9 Distributive laws For all $a, b, c \in R$, multiplication is left distributive over addition in R:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c),$$

and multiplication is right distributive over addition in R:

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c).$$

Furthermore, $(R, +, \cdot)$ is a **commutative ring** if the following extra axiom holds.

R10 Commutativity For all $a, b \in R$,

$$a \cdot b = b \cdot a.$$

At first sight, this looks like a lot of individual axioms to satisfy; so let us look at them collectively. First off, note that the axioms for addition (R1 to R5) are exactly the axioms required for $(R, +)$ to be an abelian group. As a result, we can use the 'Elementary consequences' (see, for example, Proposition 1.2 of Book B, Chapter 5) to deduce the following.

Lemma 1.2

Let $(R, +, \cdot)$ be a ring. Then the additive identity $0 \in R$ is unique. Furthermore, for every $a \in R$ the additive inverse $(-a)$ is unique.

Next, R6 to R8 tell us some rules about multiplication, but note that it is not the case that $(R, \cdot)$ is a group, because we are missing the axiom that guarantees the existence of multiplicative inverses. Finally, R9 shows us how to combine these two operations, in exactly the way we expect when multiplying and adding integers.

Early definitions of a ring did not include the multiplicative identity axiom, and this approach persists in some texts. In this case, the ring we have defined above is called a **ring with 1**, to distinguish it from a 'ring without 1'. However, all our rings will have a multiplicative identity, and so we can use the term 'ring' unambiguously.

Most of the rings in this module will also be commutative (that is, axiom R10 holds), but as we are about to see an example of a non-commutative ring, we will not for the moment assume this axiom.

Example 1.3 *Some familiar rings*

(a) $(\mathbb{Z}, +, \cdot)$, the set of integers $\mathbb{Z}$, with normal addition and multiplication, is a commutative ring.

This was our original 'template' to create the abstract definition of a ring. So it is appropriate that it is the first example we encounter.

To see this, first note that $(\mathbb{Z}, +)$ forms an infinite abelian group, with additive identity 0. This, therefore, immediately covers axioms R1 to R5. Similarly, R6 to R8 follow since multiplication over the integers is closed, associative, and the multiplicative identity is 1. Of course, multiplication over the integers is commutative, and so we can also verify axiom R10. This leaves only axiom R9, which follows as distributivity is a basic property of the integers.

(b) $(\mathbb{Q}, +, \cdot)$, the set of rational numbers, with normal addition and multiplication, forms a commutative ring. Similarly, we can readily check that the set of real numbers, $\mathbb{R}$, and the set of complex numbers, $\mathbb{C}$, are also both commutative rings under the same operations.

(c) $(M_{2\times 2}, +, \cdot)$, the set of 2×2 matrices over $\mathbb{R}$, with matrix addition and multiplication, is a ring but not a commutative ring.

Checking some of the axioms for this example takes a little more thought. For addition, closure follows since the sum of two 2×2 matrices is another 2×2 matrix. It can easily be checked that the addition of matrices is associative and commutative. The additive

identity is the zero matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and consequently the additive inverse of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

For multiplication, when we multiply two 2×2 matrices together we obtain another 2×2 matrix, and so closure (R6) is satisfied. It is straightforward (but not immediately obvious) to check that multiplication is associative, and we can check that the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ acts as the multiplicative identity for axiom R8.

Finally, axiom R9 can be readily verified, but it is well known that multiplication of matrices is not commutative and so R10 does not hold. For example:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \text{ but } \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

Exercise 1.1

The ring $\mathbb{Z} \times \mathbb{Z}$ consists of all ordered pairs of integers (a, b), with addition and multiplication defined by $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \cdot (c, d) = (ac, bd)$. Write down:

(a) $(2, 1) + (4, 5)$ and $(2, 1) \cdot (4, 5)$

(b) the additive identity element

(c) the additive inverse of the element $(2, 1)$

(d) the multiplicative identity element.

Exercise 1.2

Show that the following are commutative rings.

(a) The set of integers $\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$, equipped with addition and multiplication modulo n.

(b) The set

$$\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\},$$

equipped with the usual addition and multiplication of real numbers.

You may be wondering how we justify statements such as 'multiplication of integers is associative'. The properties of integers that you have known (not necessarily by name) and used since you first started doing arithmetic can be justified only by working with a formal definition of the integers. A well-known definition of the integers starts from a set of axioms known as 'Peano axioms', after Giuseppe Peano (1858–1932). These are then extended to give 'Peano arithmetic'. These axioms include the properties

such as associativity, commutativity and distributivity that we have been discussing, but in this module you are not expected to be familiar with these axioms; the approach we have used so far to justify such statements is sufficient.

In light of the examples above, you may now be wondering what objects *do not* form rings; so let us look at a couple before we proceed.

Example 1.4 *Two examples of non-rings*

(a) The set of non-negative integers, $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$, with the usual addition and multiplication.

We find that R1 and R2 are true, but then we cannot find additive inverses to satisfy R3: for example, we need -1 to be the additive inverse of 1, but it is not in the set $\mathbb{Z}^+$.

(b) The even integers, $2\mathbb{Z} = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$, with the usual addition and multiplication.

In this example, we find that this set does actually form an abelian group under addition. The problem comes when we try to find a multiplicative identity for axiom R8: there isn't one! If there was a multiplicative identity, $2k \in 2\mathbb{Z}$, say, then $(2k) \cdot (2n) = 2n$ for all $n \in \mathbb{Z}$. This, however, implies that $2k = 1$, and so $k \notin \mathbb{Z}$, which is a contradiction.

Having seen some examples and non-examples, we are ready to prove some simple properties that we should expect to find. Throughout these, you should be thinking how this is exactly what we expect in the examples of rings that we have just seen. Are the properties also true for each of our two non-examples, and if they aren't, why not?

Proposition 1.5 *Basic properties of rings*

Let $(R, +, \cdot)$ be a ring. Then:

(a) $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$

(b) $-(-a) = a$ for all $a \in R$

(c) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$ for all $a, b \in R$

(d) $(-a) \cdot (-b) = a \cdot b$ for all $a, b \in R$.

Proving statements like these is actually quite tricky: we need to be careful that at each stage we use only the axioms that define a ring, or results that we have derived from these axioms.

Proof We will prove parts (a)–(c), and leave part (d) as an exercise for you to do.

(a) Let $a \in R$, and set $b = 0 \cdot a$. We want to show that b is the additive identity. First, we know that $0 = 0 + 0$, and so $0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$, using axiom R9. Since we set $b = 0 \cdot a$, we have $b = b + b$. Therefore, using associativity:

$$0 = b + (-b) = (b + b) + (-b) = b + (b + (-b)) = b + 0 = b.$$

A similar argument can be used to prove that $a \cdot 0 = 0$.

(b) First, for any $a \in R$ we have $(-a) + a = 0$. Now, the element $(-a)$ of R has a unique additive inverse (by Lemma 1.2), namely $-(-a)$, from which it follows that $a = -(-a)$.

(c) We have $a \cdot b + (-a) \cdot b = (a + (-a)) \cdot b = 0 \cdot b = 0$, the last step following by part (a) of the proposition. Therefore, $(-a) \cdot b$ must be the additive inverse of $a \cdot b$, and so we have $(-a) \cdot b = -(a \cdot b)$. Similarly, we can show that $-(a \cdot b) = a \cdot (-b)$. ■

Having established these basic properties, we will relax our notation a little. We have so far used the minus sign '$-$' to denote the additive inverse of an element: $a + (-a) = 0$. When working with the integers, we are familiar with using the minus sign as a binary operation, that is, we write expressions such as '$a - b$'. We will extend this notation here. Formally, in any ring R with $a, b \in R$, we define $a - b$ to mean the result of adding the additive inverse of b to a. Put simply, $a - b = a + (-b)$.

When the context is clear, we will use two further notational simplifications, similar to those made at the start of Book B. First, we will write the juxtaposition ab to mean the product $a \cdot b$. Second, we will often refer to $(R, +, \cdot)$ by the set symbol R when it is clear which operations are being used.

Exercise 1.3

Prove part (d) of Proposition 1.5.

Exercise 1.4

What can you say about a ring in which the additive and multiplicative identities are the same, that is, a ring where $0 = 1$?

1.1 Subrings

In Exercise 1.2, we showed that the sets $\mathbb{Z}_n$ and $\mathbb{Z}[\sqrt{2}]$ were rings. In the solution, we appealed to properties of addition and multiplication of a larger set to prove some of the axioms. For example, $\mathbb{Z}[\sqrt{2}]$ is a subset of the ring $\mathbb{R}$, and hence, since we are using the same definition of addition and multiplication for both, some of the axioms hold immediately. In this subsection, we want to formalise this approach.

Definition 1.6 *Subring*

Let R be a ring. We say that S is a **subring** of R if:

(a) the set S is a subset of R

(b) S is a ring when equipped with the same binary operations as R

(c) the ring S has the same multiplicative identity as R.

Why do we insist that S has the same multiplicative identity? As we will see shortly in Worked Exercise 1.9(b), it is possible for a subset of a ring to be a ring itself, but with a different multiplicative identity. We want to be able to distinguish between this case and the definition given above.

Why, then, did we not also insist that S should have the same additive identity as R? We can deduce this from the definition given above.

Lemma 1.7

Let R be a ring, and S a subring of R. Then the additive identity of S is the same as the additive identity of R.

Proof Suppose 0_S is the additive identity in S, and 0_R that for R. Then for any $s \in S$, we have $s + 0_S = s$ as an equation in S, and so also as an equation in R. Working in R, we now have:

$$\begin{aligned} 0_R &= (-s) + s \\ &= (-s) + (s + 0_S) \\ &= ((-s) + s) + 0_S = 0_R + 0_S \\ &= 0_S. \end{aligned}$$ ■

Some obvious examples are the integers $\mathbb{Z}$ as a subring of the rationals $\mathbb{Q}$, and the rationals as a subring of the reals $\mathbb{R}$, and the reals as a subring of the complex numbers $\mathbb{C}$. We saw in Example 1.3 that all of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are rings, and for each the additive identity is 0 and the multiplicative identity is 1.

Now, as we wanted, we can state and prove a lemma justifying why we do not need to check all the axioms R1–R9 when verifying that a subset of a ring is a subring.

Lemma 1.8 *Subring criterion*

Let R be a ring and let S be a subset of R. Then S is a subring of R if:

SR1 for all $s, t \in S$, $s - t \in S$

SR2 for all $s, t \in S$, $st \in S$

SR3 the multiplicative identity of R is in S.

Axiom SR1 can be thought of as 'closure under subtraction'.

Proof Since S is a subset of R, we just need to check that S is a ring with the same binary operations and multiplicative identity as R.

Let $s \in S$. (Note that S is non-empty by SR3; so such an s exists.) By SR1, $s - s = 0 \in S$ and so $0 - s = (-s) \in S$. In addition, if $t \in S$ then $(-t) \in S$, so that $s + t = s - (-t) \in S$. This shows that $(S, +)$ is an abelian subgroup of $(R, +)$.

Multiplication is closed in S by SR2, and axiom SR3 ensures that the multiplicative identity of R is in S, and in fact it must be a multiplicative identity for S. The associative and distributive laws hold since they hold in R. Thus we have shown that S is a ring, and since the multiplicative identity coincides, it follows immediately that S is a subring of R. ■

Worked Exercise 1.9

(a) Show that the set of **Gaussian integers**, $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$ where $i = \sqrt{-1}$, equipped with the usual addition and multiplication of complex numbers, is a ring.

(b) Why is $S = \{0, 5\}$ not a subring of $\mathbb{Z}_{10}$?

Solution

(a) Since $\mathbb{Z}[i]$ is a subset of the complex numbers $\mathbb{C}$, it will be enough to prove that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. We check the conditions of Lemma 1.8. Consider two elements $a + ib$ and $c + id$ of $\mathbb{Z}[i]$. Then

SR1: $(a + ib) - (c + id) = (a - c) + i(b - d) \in \mathbb{Z}[i]$.

SR2: $(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc) \in \mathbb{Z}[i]$.

SR3: The multiplicative identity of $\mathbb{C}$ and $\mathbb{Z}[i]$ is $1 = 1 + 0i$.

It follows from the subring criterion that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, and therefore that $\mathbb{Z}[i]$ is a ring.

(b) Condition SR3 does not hold for S: the multiplicative identity of $\mathbb{Z}_{10}$ is 1, which is not in S.

Note, however, that S is a ring with the same operations as $\mathbb{Z}_{10}$. The multiplicative identity of S is 5.

The requirement that the multiplicative identity of a subring is always the same as that of the original ring has some interesting consequences. For example, the only subring of $\mathbb{Z}$ is $\mathbb{Z}$ itself, and the only subring of $\mathbb{Z}_n$ is $\mathbb{Z}_n$.

Exercise 1.5

By considering a suitable larger ring and using Lemma 1.8, prove that the following are rings:

(a) $\mathbb{Z}[\omega] = \{a + b\omega + c\omega^2 : a, b, c \in \mathbb{Z}\}$ where $\omega = e^{2i\pi/3}$, so that $\omega^3 = 1$

(b) $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$.

1.2 Zero divisors and units

The axioms that define a ring leave some opportunities for results that we might not expect. By considering the integers $\mathbb{Z}$, we already know that elements do not need to have multiplicative inverses, but in some rings even stranger things can happen.

Suppose that in $\mathbb{Z}$ we wish to solve the quadratic equation

$$x^2 - 5x + 6 = 0.$$

Factorising the left-hand side, we obtain

$$(x - 2)(x - 3) = 0.$$

It follows that either $x - 2 = 0$ or $x - 3 = 0$, and hence that $x = 2$ or $x = 3$. This method works since if in $\mathbb{Z}$ the product of two terms is 0, then one of the two terms must itself be 0.

More generally, if we know that the product of two integers a and b is 0, then we can deduce that at least one of a or b must be 0. Put another way, if we multiply two non-zero integers then their product must also be non-zero.

However, in the ring $\mathbb{Z}_6$ we have $2 \cdot 3 \equiv 0 \,(\mathrm{mod}\, 6)$. In other words, there are two non-zero elements whose product is zero. We want to be able to distinguish between elements where this does happen and those where it does not, and so we make the following definition.

Definition 1.10 *Zero divisor*

Let R be a ring. A non-zero element $a \in R$ is a **zero divisor** in R if there is a non-zero element $b \in R$ for which $ab = 0$ or $ba = 0$.

In a sense, 0 is also a zero divisor since $0 \cdot 1 = 0$ is true in every ring, but we have deliberately excluded this so that in future we do not have to keep using phrases such as 'no non-zero zero divisors'. Commutative rings with no zero divisors have a special name. Since they have several of the properties of the ring of integers, we call them **integral domains**.

We will learn more about integral domains in Chapter 12.

Example 1.11 *Rings with no zero divisors*

(a) The ring $\mathbb{Z}$ has no zero divisors – this is the fact that we used to solve the quadratic equation above. Similarly, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ also have no zero divisors.

(b) $\mathbb{Z}_7$ has no zero divisors. If $ab \equiv 0 \pmod 7$, then 7 divides ab, and hence (since 7 is prime) divides a or b (or both) in $\mathbb{Z}_7$. Thus $a = 0$ or $b = 0$.

Example 1.12 *Rings with zero divisors*

(a) $\mathbb{Z}_6$, as we saw above, does have zero divisors. Since $2 \cdot 3 \equiv 0 \pmod 6$ and $3 \cdot 4 \equiv 0 \pmod 6$, we see that all of 2, 3 and 4 are zero divisors.

However, 1 and 5 are not zero divisors since there are no numbers a and b (other than 0) in $\mathbb{Z}_6$ for which $1 \cdot a \equiv 0 \pmod 6$ or $5 \cdot b \equiv 0 \pmod 6$.

(b) The ring $M_{2\times 2}$ of 2×2 matrices with entries from $\mathbb{R}$ has zero divisors. For example, the product

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

shows that both $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ are zero divisors.

Exercise 1.6

Let R be the ring $\mathbb{Z} \times \mathbb{Z}$ (as defined in Exercise 1.1). For $a, b, c, d \in \mathbb{Z}$, addition is defined by $(a, b) + (c, d) = (a + c, b + d)$, and multiplication is given by $(a, b) \cdot (c, d) = (ac, bd)$.

(a) Give an example of a zero divisor in R.

(b) Describe all the zero divisors in R.

Now that we are aware of the possibility that rings have zero divisors, we turn again to the issue of multiplicative inverses: which elements of a ring can be multiplied together to give 1? For example, in $\mathbb{Z}$, we have $(-1) \cdot (-1) = 1$, and so -1 is its own multiplicative inverse, but there is no multiplicative inverse of 2 in $\mathbb{Z}$.

So far, we have allowed our rings to be non-commutative, that is, axiom R10 has not needed to be satisfied. Our main example of a non-commutative ring has been $M_{2\times 2}$, the ring of 2×2 matrices with entries from $\mathbb{R}$. As we move our focus to consider elements that have multiplicative inverses, the existence of non-commutative rings leads to some awkward possibilities. For example, could we have a non-commutative ring R and $a, b \in R$, with $a \cdot b = 1$, but $b \cdot a \neq 1$?

The answer to this question is 'yes' for some quite peculiar rings, but we will not explore this here.

Convention: commutative rings

To avoid the complications described above, all rings we will consider from now on will be commutative. Additionally, we will drop the word 'commutative' and simply use the term 'ring'.

With this settled, we can define a term for elements that have multiplicative inverses.

Definition 1.13 *Unit*

An element a of a (commutative) ring R is a **unit** if there is an element $a^{-1} \in R$ for which $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

The term 'unit' is overused in mathematics generally. It should be clear from the context which meaning is appropriate.

For example, in the ring $\mathbb{Z}$ the only units are 1 and -1; in the ring $\mathbb{Z}_{10}$ the units are 1, 3, 7 and 9, since

$$1 \times 1 \equiv 1 \,(\text{mod } 10),\ 3 \times 7 \equiv 1 \,(\text{mod } 10) \text{ and } 9 \times 9 \equiv 1 \,(\text{mod } 10).$$

Exercise 1.7

List all the units in the following rings.

(a) $\mathbb{Z}_{12}$

(b) $\mathbb{Z} \times \mathbb{Z}$

(c) $\mathbb{Z}[i]$

Although in the above exercise it was relatively straightforward to list the units in specific rings, this is not always the case. We will see examples in Chapter 12 where finding units is rather more difficult.

Exercise 1.8

(a) Write down all the zero divisors in the rings $\mathbb{Z}_n$, for $n = 8$, 9, 10, 11 and 12.

(b) Give a description of all the zero divisors in $\mathbb{Z}_n$, for any given natural number n.

(c) Which rings $\mathbb{Z}_n$ have no zero divisors?

(d) Show that in a ring R, if a is a unit then it cannot be a zero divisor.

1.3 Fields

In the definition of a ring, the axiom we were missing for the non-zero elements to form a group under multiplication was that every element has an inverse: in other words, every non-zero element is a unit. A ring with this property is called a *field*.

Definition 1.14 *Field*

A (commutative) ring R is a **field** if the additive and multiplicative identities are distinct, and every non-zero element is a unit.

In other words, R has distinct elements 0 and 1, axioms R1–R10 hold, and so does:

R11 Multiplicative inverses For every non-zero $a \in R$, there exists $a^{-1} \in R$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Recall Exercise 1.4, which tells us what happens when $0 = 1$ in a ring.

In the same way that we have been using the symbol R to refer to a general ring, we will use F for a field. You may find that in other texts the letter K is used: this is because the German mathematician Richard Dedekind (1831–1916) used the term *Körper* for field, which translates as 'body'.

Example 1.15 *Some well-known fields*

The following are easily seen to be fields.

(a) $\mathbb{Q}$: Since $\frac{p}{q} \cdot \frac{q}{p} = 1$ for non-zero $p, q \in \mathbb{Z}$, the inverse of $\frac{p}{q}$ is $\frac{q}{p}$.

(b) $\mathbb{R}$: The inverse of any non-zero $x \in \mathbb{R}$ is $x^{-1} = \frac{1}{x}$.

(c) $\mathbb{C}$: The inverse of $z = a + ib \neq 0$ is

$$z^{-1} = \frac{1}{a+ib} = \left(\frac{a}{a^2+b^2}\right) + i\left(\frac{-b}{a^2+b^2}\right).$$

Example 1.16 *Some rings that are not fields*

(a) $\mathbb{Z}$ is not a field: as we observed in the previous subsection, the only units are -1 and 1, and so no other integer has a multiplicative inverse.

(b) $\mathbb{Z}_6$ is not a field: for example, the element 2 is not a unit.

Since $\mathbb{Z}_6$ has sufficiently few elements, we can verify this by testing all multiples of 2 modulo 6:

$$2 \cdot 0 \equiv 0, \quad 2 \cdot 1 \equiv 2, \quad 2 \cdot 2 \equiv 4,$$
$$2 \cdot 3 \equiv 0, \quad 2 \cdot 4 \equiv 2, \quad 2 \cdot 5 \equiv 4.$$

Exercise 1.9 *Properties of multiplicative inverses*

Let F be a field, and $a, b \in F$ non-zero elements. Prove the following.

(a) $(a^{-1})^{-1} = a$

(b) $(ab)^{-1} = b^{-1}a^{-1}$

The second example of a non-field, $\mathbb{Z}_6$, raises an interesting point: we used the element 2, which is a zero divisor, since $2 \cdot 3 \equiv 0 \pmod 6$. In Exercise 1.8(d), we showed that any element that is a unit cannot also be a zero divisor. As a consequence of this, we have the following lemma.

Lemma 1.17

Let F be a field. Then F has no zero divisors.

Proof Suppose $a \neq 0$ is a zero divisor in F, and so there exists a non-zero $b \in F$ such that $a \cdot b = 0$. By axiom R11, there exists $b^{-1} \in F$ such that $b \cdot b^{-1} = 1$. Now

$$a = a \cdot 1 = a \cdot (b \cdot b^{-1}) = (a \cdot b) \cdot b^{-1} = 0 \cdot b^{-1} = 0,$$

which is a contradiction. ■

In Exercise 1.8, we stated that $\mathbb{Z}_n$ has no zero divisors if, and only if, n is a prime number. In fact, each element of $\mathbb{Z}_n$ is either a zero divisor or a unit.

Lemma 1.18

Let $a \in \mathbb{Z}_n$ be non-zero, where $n \in \mathbb{N}$. Then a is either a zero divisor or a unit.

Proof If $\text{hcf}(a, n) = 1$, then by Proposition 4.7 of Book A, Chapter 1 there exist $s, t \in \mathbb{Z}$ such that $sa + tn = 1$. Therefore, $sa \equiv 1 \pmod n$, and hence a is a unit in $\mathbb{Z}_n$.

On the other hand, if $\text{hcf}(a, n) = d > 1$, then $a = dh$ and $n = dk$ for some h, k, with $1 < k < n$ and $\text{hcf}(h, k) = 1$. Then

$$ak = (dh)k = (hd)k = h(dk) = hn \equiv 0 \pmod n,$$

which proves that a is a zero divisor. ■

For any non-zero element a, one of two things can happen: either $\text{hcf}(a, n) = 1$, or $\text{hcf}(a, n) > 1$. We consider each of these possibilities in turn.

This lemma immediately gives us an important consequence.

Corollary 1.19

The ring $\mathbb{Z}_n$ is a field if, and only if, n is a prime number.

The fields $\mathbb{Z}_p$, where p is a prime number, contain only finitely many elements, and are therefore called **finite fields**. We will not explore finite fields further in this chapter, as we will study them in some detail in Book E, Chapter 19. However, we will make two quick remarks.

First, the fields $\mathbb{Z}_p$, for p prime, are not the only finite fields: in fact, there exists a finite field with q elements if, and only if, q is a power of a prime number.

Second, if we repeatedly add the multiplicative identity to itself, $1 + 1 + \cdots + 1$, we will at some point end up with the additive identity. For example, in $\mathbb{Z}_3$, $1 + 1 + 1 = 3 \equiv 0 \pmod 3$. The number of times we add 1 to itself to get 0 is known as the **characteristic**. So, the characteristic of $\mathbb{Z}_3$ is 3.

We finish this section by defining subfields, and developing an analogue to the subring criterion (Lemma 1.8) that allows us to avoid checking all the axioms.

Definition 1.20 *Subfield*

A subset S of a field F that is a field with the same binary operations and multiplicative identity as F is known as a **subfield** of F.

Lemma 1.21 *Subfield criterion*

Let F be a field, and let S be any subset of F. Then S is a subfield of F if:

(a) S is a subring of F

(b) every non-zero element of S has a multiplicative inverse in S.

Consequently, to determine whether S is a subfield of F or not, we need to check only axioms SR1–SR3 from Lemma 1.8 and axiom R11 from Definition 1.14.

Proof Since S is a subset of F, we just need to check that S is a field with the same binary operations and identities as F. First, condition (a) tells us that S has the same multiplicative identity as F, and that all the axioms R1–R10 hold. This leaves only axiom R11 about the existence of multiplicative inverses, but this is guaranteed by condition (b). ■

Exercise 1.10

By considering suitable larger fields, show that the following are fields.

(a) $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$

(b) $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$

2 Polynomials over fields

As we saw in Subsection 1.2, rings can have unexpected properties. To avoid some of these issues, we are going to spend the remainder of this chapter studying polynomial rings where the coefficients of the polynomials come from a field, as these are generally much better behaved than generic rings. There will be one exception to this rule: towards the end of Section 4 we will encounter the polynomial ring $\mathbb{Z}[x]$, even though $\mathbb{Z}$ is not a field.

2.1 Polynomial rings

Definition 2.1 *Polynomial ring over a field*

Let F be a field. Then a **polynomial over** F with the variable x is a polynomial of the form $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, where $a_0, a_1, a_2, \ldots, a_n \in F$ and $n \geq 0$, where x^0 is defined to be 1.

The **polynomial ring over** F is

$$F[x] = \{a_0 + a_1x + \cdots + a_nx^n : a_0, a_1, \ldots, a_n \in F, n \geq 0\}.$$

We have seen the square bracket notation '$F[x]$' a few times already in this chapter, and it is worth verifying that our use of it has been consistent. For example, in Exercise 1.10(b) we encountered $\mathbb{Q}[\sqrt{2}]$, which was defined to be the set $\{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. The above definition would define it as $\{a_0 + a_1\sqrt{2} + \cdots + a_n(\sqrt{2})^n : a_i \in \mathbb{Q}\}$, but of course $(\sqrt{2})^2 = 2$ is already an element of $\mathbb{Q}$, and so the higher terms of $\sqrt{2}$ can simply be included in the coefficients a_0 and a_1.

To be clear that $F[x]$ is indeed a ring, we need to define 'addition' and 'multiplication' within $F[x]$, and then be sure that all the axioms hold. How we add and multiply polynomials should come as no surprise: first, we need to remember that the field F came with its own addition and multiplication, and use these to build the natural definitions. If in doubt, think about how it is done in $\mathbb{R}[x]$, that is, when the coefficients are all real numbers.

Addition is defined by

$$\left(\sum_{i=0}^{n} a_ix^i\right) + \left(\sum_{i=0}^{m} b_ix^i\right) = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i.$$

Note that the sum on the right-hand side of this equation goes up to $\max(m, n)$: so, if $n < m$ then we add 'dummy' coefficients $a_{n+1} = a_{n+2} = \cdots = a_m = 0$.

What is the zero of $F[x]$? If '0' is the symbol for the zero in the field F, then let 0_x be the zero in $F[x]$. It is defined as the polynomial where every coefficient is equal to 0; so in fact we have $0_x = 0$. However, we will

continue to use the symbol 0_x to highlight when we are working in the polynomial ring $F[x]$.

We can now work out additive inverses:

$$\left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} (-a_i) x^i\right) = \sum_{i=0}^{n} \left(a_i + (-a_i)\right) x^i = 0_x.$$

The closure, associativity and commutativity of addition in $F[x]$ follow quite easily: remember that we can assume they hold in F.

Next, we define multiplication:

$$\left(\sum_{i=0}^{n} a_i x^i\right) \cdot \left(\sum_{j=0}^{m} b_j x^j\right) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \cdot b_j\right) x^k.$$

At first sight, this definition looks rather offputting. So, let us quickly consider an example to reassure ourselves that this is exactly what we expect to see. Take polynomials $f(x) = 1 + 5x - 4x^2$ and $g(x) = 2 + 3x$ over $\mathbb{R}$. Then

$$\begin{aligned} f(x)g(x) &= (1 + 5x - 4x^2) \cdot (2 + 3x) \\ &= \big(1 \cdot 2\big) + \big(1 \cdot 3 + 5 \cdot 2\big)x + \big(5 \cdot 3 + (-4) \cdot 2\big)x^2 + \big((-4) \cdot 3\big)x^3 \\ &= 2 + 13x + 7x^2 - 12x^3. \end{aligned}$$

What is the multiplicative identity of $F[x]$? As we did with the zero, we will use the special symbol 1_x to distinguish it from the '1' in F, even though $1_x = 1$.

The remaining axioms R6–R9 are readily verified, as a consequence of the axioms for the field F. In addition, since multiplication in F is commutative, it follows that $F[x]$ is a commutative ring, that is, axiom R10 holds.

Example 2.2 *Some common polynomial rings*

The polynomial ring $\mathbb{Q}[x]$ will appear in Section 4.

(a) The set of polynomials with rational coefficients, $\mathbb{Q}[x]$, with $0_x = 0$ and $1_x = 1$, is a ring since $\mathbb{Q}$ is a field.

(b) $\mathbb{C}[x]$ represents the polynomials with complex coefficients, and $\mathbb{R}[x]$ the polynomials with real coefficients. Note that $\mathbb{Q}[x] \subset \mathbb{R}[x] \subset \mathbb{C}[x]$.

(c) The polynomial ring $\mathbb{Z}_2[x]$ consists of all polynomials whose coefficients are all either 0 or 1. More generally, we can consider the polynomial ring $\mathbb{Z}_p[x]$ where p is a prime number.

Worked Exercise 2.3

Let $f(x) = 1 + 2x$ and $g(x) = 1 + 3x + 2x^3$ be polynomials in $\mathbb{Z}_5[x]$, the ring of polynomials over the field $\mathbb{Z}_5$, equipped with addition and multiplication modulo 5.

(a) Calculate and simplify (i) $f(x) + g(x)$, and (ii) $3f(x) - g(x)$.

(b) Compute $f(x)g(x)$.

Solution

(a) (i) $f(x) + g(x) = (1 + 1) + (2 + 3)x + 2x^3 \equiv 2 + 2x^3 \pmod 5$

(ii)
$$\begin{aligned} 3f(x) - g(x) &= 3(1 + 2x) - (1 + 3x + 2x^3) \\ &= (3 + (3 \times 2)x) - 1 - 3x - 2x^3 \\ &\equiv (3 + x) + 4 + 2x + 3x^3 \\ &= (3 + 4) + (1 + 2)x + 3x^3 \\ &\equiv 2 + 3x + 3x^3 \pmod 5 \end{aligned}$$

(b)
$$\begin{aligned} f(x)g(x) &= (1 + 2x)(1 + 3x + 2x^3) \\ &= 1 + 5x + 6x^2 + 2x^3 + 4x^4 \\ &\equiv 1 + x^2 + 2x^3 + 4x^4 \pmod 5 \end{aligned}$$

Note that it is sometimes easier when working in $\mathbb{Z}_p[x]$ to calculate the polynomials over $\mathbb{Z}$, and then reduce the coefficients modulo p at the end.

Exercise 2.1

Let $f(x) = 2x^3 - 3$ and $g(x) = 9x^4 + 7x^2 - x + 1$. Compute (i) $f(x) + g(x)$ and (ii) $f(x) \cdot g(x)$ in the following polynomial rings:

(a) $\mathbb{Q}[x]$

(b) $\mathbb{Z}_7[x]$

(c) $\mathbb{Z}_2[x]$.

2.2 The degree of a polynomial

Definition 2.4 *Degree of a polynomial*

Let F be a field, and let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial in $F[x]$. The **degree** of $f(x)$, $\deg(f)$, is the largest $k \geq 0$ for which $a_k \neq 0$.

This non-zero coefficient a_k is known as the **leading coefficient**.

At this point, it is also worth making a remark about notation: strictly speaking, we should have written $\deg(f(x))$ rather than $\deg(f)$, but this is rather clumsy. For the same reason, in future we will occasionally drop the '(x)' from other equations involving polynomials, where the meaning is still sufficiently clear.

Example 2.5 *The degree of a polynomial*

In Exercise 2.1, the polynomial $f(x) = 2x^3 - 3$ has degree 3 in $\mathbb{Q}[x]$ and $\mathbb{Z}_7[x]$, but $f(x) \equiv 1 \pmod 2$ and so $f(x)$ has degree 0 in $\mathbb{Z}_2[x]$.

On the other hand, the polynomial $g(x) = 9x^4 + 7x^2 - x + 1$ has degree 4 in all of $\mathbb{Q}[x]$, $\mathbb{Z}_7[x]$ and $\mathbb{Z}_2[x]$.

Convention: degree of the zero polynomial

By convention, the degree of the polynomial 0_x will be $-\infty$, which should be interpreted as meaning we can choose the degree of 0_x to be as largely negative as we need.

This choice is made so that the following lemma is consistent when one or both of the polynomials f and g are equal to 0_x.

Lemma 2.6

Let F be a field, and let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Then $\deg(fg) = \deg(f) + \deg(g)$.

This seemingly innocuous lemma actually turns out to be quite important, as we will see later in this chapter and the next. Before we prove it, let us pause to think what would happen if we replaced the field F with some arbitrary ring R. We consider an example in the following exercise.

Exercise 2.2

Consider the polynomials $f(x) = 2x^2 + x + 3$ and $g(x) = 3x + 1$ in the polynomial ring $\mathbb{Z}_6[x]$. Find:

(a) $\deg(f)$

(b) $\deg(g)$

(c) $\deg(fg)$.

What went wrong? We expected the degree of $f(x)g(x)$ to be 3, but the coefficient of x^3 disappeared because $2 \cdot 3 \equiv 0 \pmod 6$. This, of course, is because 2 and 3 are zero divisors in $\mathbb{Z}_6$, and we know from Lemma 1.17 that fields cannot have zero divisors.

Proof of Lemma 2.6 First, if either $f(x)$ or $g(x)$ is equal to 0_x, then $fg = 0_x$ and $\deg(fg) = -\infty$ as we should expect. So we now suppose that $f(x) \neq 0_x$ and $g(x) \neq 0_x$.

Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$ be polynomials over F, where $\deg(f) = n$ and $\deg(g) = m$. In particular, this means that $a_n \neq 0$ and $b_m \neq 0$. Now, from the definition of multiplication in $F[x]$,

$$f(x)g(x) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \cdot b_j \right) x^k.$$

The highest term in this sum is $a_nb_mx^{m+n}$, and so we certainly have $\deg(fg) \leq \deg(f) + \deg(g)$. Moreover, since $a_n, b_m \in F$ are both not equal to zero, it follows by Lemma 1.17 that their product a_nb_m is also not 0. Thus $\deg(fg) = m + n = \deg(f) + \deg(g)$. ■

Meanwhile, what happens to the degree if we add two polynomials? Clearly, from the definition, $\deg(f + g) \leq \max(\deg(f), \deg(g))$, but can we replace the '$\leq$' with '$=$'? Again, let us first consider a couple of examples.

Exercise 2.3

Calculate the degree of $f(x) + g(x)$ over the stated ring:

(a) $f(x) = 2x^3 + 4x^2 - 1$ and $g(x) = 3x^3 + 3x^2 - 2x$ over $\mathbb{Z}_5$

(b) $f(x) = 1 + x$ and $g(x) = 1 - x$ over $\mathbb{Q}$.

In the above exercise, you should have found that it is indeed possible for the degree of the sum of two polynomials to be less than the sum of the degrees. In each case, the top term of $f + g$ disappeared because $\deg(f) = \deg(g)$, and the leading coefficients of f and g were additive inverses of each other.

Lemma 2.7

Let F be a field, and let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Then $\deg(f + g) \leq \max(\deg(f), \deg(g))$.

Proof This follows directly from the definition of addition for polynomials in $F[x]$. ■

2.3 Basic properties of polynomial rings over fields

Now we have defined and developed the concept of the degree of a polynomial, we are in a position to prove some results to reassure us that polynomials over fields behave in the way that we expect them to.

First, the property that the field F has no zero divisors can be shown to imply that there are also no zero divisors in $F[x]$.

Lemma 2.8

Let F be a field, and let $f(x), g(x) \in F[x]$ be polynomials such that $f(x)g(x) = 0_x$. Then either $f(x) = 0_x$, or $g(x) = 0_x$ (or both).

Proof Since by Lemma 2.6 we have $-\infty = \deg(0_x) = \deg(fg)$ $= \deg(f) + \deg(g)$, it follows that at least one of $f(x)$ or $g(x)$ must have degree $-\infty$, that is, $f(x) = 0_x$ or $g(x) = 0_x$. ■

We might now wonder whether in fact $F[x]$ is a field: is every element a unit? This turns out not to be the case.

Lemma 2.9

An element of a polynomial ring $F[x]$ over a field F has a multiplicative inverse (that is, the element is a unit) if, and only if, it has degree 0.

Proof First, a polynomial $f(x)$ of degree 0 is simply a non-zero element of the field F: $f(x) = a_0$ with $a_0 \in F$. Since F is a field, there exists $a_0^{-1} \in F$ such that $a_0 \cdot a_0^{-1} = 1$. Therefore, the inverse of $f(x)$ in $F[x]$ is $f^{-1}(x) = a_0^{-1}$.

Conversely, for a polynomial $f(x) \in F[x]$, suppose that $g(x) \in F[x]$ is the multiplicative inverse. Then $f(x)g(x) = 1_x$, and so by Lemma 2.6 $\deg(f) + \deg(g) = \deg(1_x) = 0$. By definition, the degree of a polynomial is either $-\infty$ or an integer value greater than or equal to 0. So the only possible situation in which $\deg(f) + \deg(g) = 0$ arises is when f has degree 0. ■

Since every unit in a polynomial ring $F[x]$ is actually in F, we will typically refer to $u \in F[x]$ rather than $u(x) \in F[x]$ from now on.

Lemma 2.9 implies that $F[x]$ is not itself a field, since we cannot find inverses for anything with degree ≥ 1. However, despite this lack of inverses in $F[x]$, we can still cancel terms from equations.

Lemma 2.10 *Cancellation of polynomials*

If $f(x)$ is a non-zero polynomial and $f(x)g(x) = f(x)h(x)$ in $F[x]$, then $g(x) = h(x)$.

Proof We can rearrange the equation $fg = fh$ to obtain $fg - fh = 0$, which by distributivity gives us $f(g - h) = 0$. Now $g - h$ is a polynomial over F, so by Lemma 2.8 either $f(x) = 0_x$ (which by assumption is not true), or $g(x) - h(x) = 0_x$. Therefore, we must have $g(x) = h(x)$. ■

We finish this section with a couple of definitions that we will need later in this chapter.

Definition 2.11 *Associate and monic*

Let F be a field, and $f(x), g(x) \in F[x]$.

(a) We say that $f(x)$ is an **associate** of $g(x)$ if $f(x) = ug(x)$ where u is a unit in $F[x]$.

(b) The polynomial $f(x)$ is **monic** if the leading coefficient is equal to 1.

We will generalise the concept of 'associate' to other rings in Chapter 12.

Exercise 2.4

Let F be a field, and $f(x)$ a non-zero polynomial in $F[x]$. Prove the following.

(a) If $g(x) \in F[x]$ is an associate of $f(x)$, then $\deg(g) = \deg(f)$.

(b) There exists a unique monic polynomial that is an associate of $f(x)$.

3 Divisibility of polynomials

This section seeks to develop some results concerning the divisibility of polynomials that mirror results in Book A, Chapter 1 for the integers.

3.1 Polynomial division

You may already have had practice at dividing one polynomial over the real numbers by another and calculating the remainder. Here, we want to extend this 'long division' to polynomials over an arbitrary field F. Before we do this, however, we should justify that what we are doing is valid, and can be done in a unique way: what we need is a division algorithm for polynomials, to mirror Theorem 4.1 in Book A, Chapter 1. In fact, as we proceed through this section you may like periodically to glance back at Section 4 of Chapter 1, and observe the similarities.

Theorem 3.1 *Division Algorithm for polynomials*

Let F be a field and let f and g be polynomials in $F[x]$, with $g \neq 0_x$. Then there exist unique polynomials q and r in $F[x]$ such that

(a) $\deg(r) < \deg(g)$, and

(b) $f = qg + r$.

The polynomials q and r are respectively known as the **quotient** and **remainder** when dividing f by g.

There are two parts to this proof. First, we prove the existence of a suitable q and r using induction, and then we prove that they are unique.

Proof If $f = 0_x$ or $\deg(f) < \deg(g)$, then we can simply take $q = 0_x$ and $r = f(x)$. Thus from now on we will assume $\deg(f) \geq \deg(g)$. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$, with $a_n \neq 0$ and $b_m \neq 0$ so that $\deg(f) = n$ and $\deg(g) = m$. We use induction on $d = n - m \geq 0$ to prove the existence of a suitable $q(x)$ and $r(x)$.

Remember F is a field, and so we can find b_m^{-1} such that $b_m b_m^{-1} = 1$.

The base case is $d = 0$, when $m = n$. In this case, we can take $q = a_n b_m^{-1}$, and $r = f - qg$. Note that $\deg(r) < \deg(g)$ because the coefficient of x^m in r disappears: $a_m - a_m b_m^{-1} b_m = 0$.

We have expressly constructed f_1 so that $\deg(f_1) < \deg(f)$.

For the inductive step, suppose that the statement is true for $d = 0, 1, \ldots, k-1$. Now consider $d = k$. Define $f_1 = f - a_n b_m^{-1} x^k g$. There are three cases to consider: $f_1 = 0_x$, $\deg(f_1) < \deg(g)$ and $\deg(f_1) \geq \deg(g)$. If $f_1 = 0_x$, then we take $q = a_n b_m^{-1} x^{n-m}$ and $r = 0_x$. If $\deg(f_1) < \deg(g)$, then we take $q = a_n b_m^{-1} x^{n-m}$ and $r = f_1$. Thus we now suppose $\deg(f_1) \geq \deg(g)$. By induction, we can find $q_1(x), r_1(x)$ such that $f_1 = q_1 g + r_1$, with $\deg(r_1) < \deg(g)$. We now substitute back:

$$\begin{aligned} f &= a_n b_m^{-1} x^{n-m} g + f_1 \\ &= a_n b_m^{-1} x^{n-m} g + (q_1 g + r_1) \\ &= \left(a_n b_m^{-1} x^{n-m} + q_1\right) g + r_1 \end{aligned}$$

and so we take $q = a_n b_m^{-1} x^{n-m} + q_1$ and $r = r_1$.

For uniqueness of q and r, suppose that we can find another pair of polynomials q^*, r^* with $f = q^* g + r^*$. Then we have $qg + r = q^* g + r^*$, which can be rearranged to give $(r - r^*) = (q^* - q)g$. Applying Lemma 2.6, we get $\deg(r - r^*) = \deg(q^* - q) + \deg(g)$, but since $\deg(g) > \deg(r - r^*)$ this is a contradiction unless $r = r^*$ and $q^* = q$. ■

Actually finding q and r, when working in a polynomial ring, requires long division of polynomials. We will work through the steps of one example, in case you are not familiar with the process or need reminding.

Worked Exercise 3.2

Find the quotient and remainder on dividing $f(x) = x^4 + x^2 + x - 2$ by $g(x) = x^2 - 1$ in $\mathbb{Q}[x]$.

Solution

We want to find $q(x)$ and $r(x)$ with $f(x) = g(x)q(x) + r(x)$ and the degree of $r(x)$ smaller than that of $g(x)$ (so that $q(x)$ has as high a degree as possible).

First, we write the terms of $f(x)$ and $g(x)$ in descending powers of x, leaving a 'gap' for any powers of x that are missing in $f(x)$.

$$\begin{array}{r|l} x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Next, take the highest term of $f(x)$, which is x^4, and divide it by the highest term of $g(x)$, which is x^2. The result is the first term, x^2, of the quotient $q(x)$, which we write above the line.

$$\begin{array}{r|l} & x^2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Now, multiply $g(x)$ by this first term of $q(x)$, and write the terms below the corresponding terms of $f(x)$. Then subtract these terms from $f(x)$.

$$\begin{array}{r|l} & x^2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

We repeat the process with this new polynomial, $2x^2 + x - 2$, in place of $f(x)$: take the highest term of the new polynomial, which is $2x^2$, and divide it by the highest term of $g(x)$, x^2, to get the next term, 2, of $q(x)$.

$$\begin{array}{r|l} & x^2 \quad + \quad 2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Then multiply $g(x)$ by this new term of $q(x)$, and subtract from the polynomial at the bottom.

$$\begin{array}{r|l} & x^2 \quad + \quad 2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \\ & \qquad 2x^2 \qquad \quad - \quad 2 \\ \hline & \qquad\qquad x \end{array}$$

We would now repeat the process, except that the highest term of $g(x)$ is now larger than the highest term of the polynomial we want to divide, and hence what is left is the remainder. This gives $q(x) = x^2 + 2$ and $r(x) = x$.

Exercise 3.1

For each of the following polynomials $f(x), g(x)$ in $\mathbb{Q}[x]$, find the quotient $q(x)$ and the remainder $r(x)$ for the division of $f(x)$ by $g(x)$.

(a) $f(x) = x^3 - 2x^2 + 3x - 1,\ g(x) = x - 1$

(b) $f(x) = 2x^4 - x + 1,\ g(x) = x^2 + 1$

(c) $f(x) = 3x^3 - 2x^2 + 1,\ g(x) = 2x + 1$

With a division algorithm in hand, we now state precisely what we mean when we say that one polynomial divides another: the remainder in the above process must equal the zero polynomial 0_x in $F[x]$.

Definition 3.3 *Factors of a polynomial*

Let F be a field, and f, g polynomials in $F[x]$. We say that g **divides** f (or g is a **factor** of f) if there is some polynomial $h \in F[x]$, such that $f = gh$.

If f is non-zero and neither g nor h are units then g and h are **proper factors** of f.

Notice that if g and h are proper factors of f in $F[x]$, then $0 < \deg(g) < \deg(f)$ and $0 < \deg(h) < \deg(f)$: if $\deg(g) = 0$ then $g \in F$ would be a unit, while if $\deg(g) = \deg(f)$ then Lemma 2.6 of this chapter would imply $\deg(h) = 0$, and so h would be a unit.

Once again, the zero polynomial 0_x behaves slightly differently: every polynomial is a factor of 0_x since $0_x = f(x) \cdot 0_x$. The following exercise will enable you to develop properties of polynomial division that are directly analogous to those we proved for the integers in Section 4 of Book A, Chapter 1.

Exercise 3.2 *Properties of division for polynomials*

Let F be a field, and let f, g and h be non-zero polynomials in $F[x]$. Prove the following statements.

(a) If g divides f then g divides $f + gh$ for any polynomial $h \in F[x]$.

(b) If h divides g and g divides f then h divides f.

(c) If h divides f and h divides g then h divides $af + bg$ for any polynomials $a, b \in F[x]$.

(d) The polynomials f and g are associates if, and only if, g divides f and f divides g.

3.2 Highest common factors

We have now laid the foundations we need to enable us to describe the analogue of the Euclidean Algorithm for polynomials, and hence develop the concept of highest common factor. We proceed in exactly the same way as we did for the integers in Book A, Chapter 1.

Definition 3.4 *Highest common factor of two polynomials*

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then the **highest common factor** of f and g, written $\text{hcf}(f, g)$, is a monic polynomial of largest degree satisfying the following:

(a) $\text{hcf}(f, g)$ divides both f and g.

(b) Any polynomial $d \in F[x]$ that divides both f and g must also divide $\text{hcf}(f, g)$.

As with the highest common factor of two integers, the hcf of two polynomials is unique, as the next result will show. First, however, we should ask ourselves whether the condition that $\text{hcf}(f, g)$ is monic is strictly necessary. It is straightforward to see that yes, it is necessary to ensure uniqueness; otherwise, any associate of $\text{hcf}(f, g)$ would also satisfy conditions (a) and (b) of Definition 3.4.

Theorem 3.5

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then $\text{hcf}(f, g)$ is unique, and there exist polynomials $a, b \in F[x]$ such that

$$\text{hcf}(f, g) = af + bg.$$

Proof We consider the set of polynomials

$$S = \{af + bg : a, b \in F[x]\}.$$

Compare the proof of this theorem to the proof of Proposition 4.7 in Book A, Chapter 1.

First note that if h is a polynomial in S, any associate of h is also in S, since these are formed by multiplying h by an element of F. By applying the Well-Ordering Principle to the degrees of the polynomials in S, we can choose from S a polynomial of smallest degree, $h = af + bg$. Moreover, since every associate of h is also in the set S, we can assume h is monic.

We claim that this monic polynomial h is unique. If there existed in S another monic polynomial $h_1 = a_1 f + b_1 g$ of the same degree as h, then

$$\begin{aligned} h - h_1 &= af + bg - (a_1 f + b_1 g) \\ &= (a - a_1)f + (b - b_1)g \end{aligned}$$

is another polynomial from S. Since h and h_1 are both monic, it follows that either $h - h_1 \neq 0_x$ has smaller degree than h (which is impossible), or $h - h_1 = 0_x$, so that h and h_1 are identical.

To complete the proof, we need to prove that $h = \text{hcf}(f, g)$. For this, we use the Division Algorithm (Theorem 3.1) to find polynomials q and r such that $f = qh + r$ with $\deg(r) < \deg(h)$. We want to show that $r = 0_x$. To do this, we demonstrate that r is a member of S:

$$\begin{aligned} r &= f - qh \\ &= 1 \cdot f - q\,(af + bg) \\ &= (1 - qa)f + (-qb)g \in S. \end{aligned}$$

However, r has strictly smaller degree than h, and so we must have $r = 0_x$. Hence $f = qh$ and so f is a multiple of h. A similar argument tells us that g is a multiple of h, and hence h is a common factor.

Finally, we have to show that h is the *highest* of these factors. This follows by Exercise 3.2(c): for any common factor d of f and g, we know that d divides $af + bg = h$. ■

Both Definition 3.4 and Theorem 3.5 have the condition that f and g cannot both be equal to the zero polynomial 0_x of $F[x]$. When working with the integers in Book A, Chapter 1 we simply stated that $\text{hcf}(0, 0)$ does not exist, and this would seem to make sense here too; thus, we declare that $\text{hcf}(0_x, 0_x)$ does not exist.

Some authors instead define $\text{hcf}(0_x, 0_x) = 0_x$.

Following the pattern laid out by Chapter 1, we can now give meaning to the term 'coprime' in the context of polynomials.

Definition 3.6 *Coprime*

Let F be a field, and let f, g be polynomials in $F[x]$, not both equal to 0_x. If $\text{hcf}(f, g) = 1_x$ then we say that f and g are **coprime**.

As a direct consequence of Theorem 3.5, we therefore have the following result.

Corollary 3.7

The polynomials $f, g \in F[x]$, not both equal to 0_x, are coprime if, and only if, there exist polynomials $a, b \in F[x]$ such that $af + bg = 1_x$.

Exercise 3.3

Let F be a field, and f, g and h be polynomials in $F[x]$. Prove the following.

(a) If $\text{hcf}(f, g) = 1_x$ and both f and g divide h, then fg divides h.

(b) If f divides gh and $\text{hcf}(f, g) = 1_x$, then f divides h.

Part (b) is the polynomial ring equivalent to Euclid's Lemma, Theorem 4.12 of Book A, Chapter 1.

3.3 The Euclidean Algorithm

In the last subsection, we proved the existence and uniqueness of the highest common factor of two polynomials over a field. However, as was the case in Book A, Chapter 1, the proof of this fact was not hugely enlightening on how to find the highest common factor. In this subsection, we will see how to apply the Division Algorithm (Theorem 3.1) to carry out practical calculation of the highest common factor.

Strategy: Euclidean Algorithm to find the hcf of two polynomials

Given a field F, and two polynomials $f, g \in F[x]$, not both equal to 0_x, let $f^* = f$ and $g^* = g$.

1. Apply the Division Algorithm (Theorem 3.1) to f^*, g^* to find $q, r \in F[x]$ for which $f^* = qg^* + r$.
2. If $r = 0_x$ then stop: $\text{hcf}(f, g) = a^{-1}g^*$, where $a \in F$ is the coefficient of the highest power of x in g^*.
3. Otherwise, $r \neq 0_x$. Replace f^* by g^*, and g^* by r, and go back to step 1.

Worked Exercise 3.8

In $\mathbb{Q}[x]$, find $\text{hcf}(3x^4 + 2x^3 + x^2 - 4x + 1,\ x^2 + x + 1)$.

Solution

For convenience later, we will set $f(x) = 3x^4 + 2x^3 + x^2 - 4x + 1$ and $g(x) = x^2 + x + 1$, so that we are trying to find $\text{hcf}(f, g)$. First, we divide f by g using polynomial long division, to obtain

$$f(x) = (3x^2 - x - 1)g(x) + (-2x + 2).$$

The next step of the Euclidean Algorithm requires that we divide $g(x)$ by the remainder, $-2x + 2$. This gives

$$g(x) = (-\tfrac{1}{2}x - 1)(-2x + 2) + 3.$$

As has happened here, when the remainder is a constant (that is, has degree zero) we know that the next step of the algorithm must give us zero remainder, and so the following final step could actually be omitted:

$$-2x + 2 = (-\tfrac{2}{3}x + \tfrac{2}{3}) \cdot 3 + 0.$$

Therefore, $\text{hcf}(f, g) = \frac{1}{3} \cdot 3 = 1$.

Exercise 3.4

(a) Use the Euclidean Algorithm to find $\text{hcf}(x^3 + 2x^2 - x - 2, x^2 - 4x + 3)$ in $\mathbb{Q}[x]$.

(b) Hence, or otherwise, find polynomials s, t in $\mathbb{Q}[x]$ for which $x - 1 = s(x^3 + 2x^2 - x - 2) + t(x^2 - 4x + 3)$.

As well as providing a method to find the hcf of two polynomials, the above exercise shows that the Euclidean Algorithm also helps us to find the polynomials $a(x), b(x)$ such that $\text{hcf}(f, g) = af + bg$ (as was guaranteed by Theorem 3.5). This was a relatively straightforward task in the above exercise, but for more complicated scenarios the strategy is to work backwards through the Euclidean Algorithm (as we did back in Book A, Chapter 1). We demonstrate this in a worked exercise.

Worked Exercise 3.9

Let $f(x) = 3x^4 + 2x^3 + x^2 - 4x + 1$ and $g(x) = x^2 + x + 1$. In $\mathbb{Q}[x]$, find polynomials $a(x)$ and $b(x)$ such that $1 = a(x)f(x) + b(x)g(x)$.

Solution

In Worked Exercise 3.8, we showed that $\text{hcf}(f, g) = 1$. Working backwards through the Euclidean Algorithm:

$$\begin{aligned} 3 &= (\tfrac{1}{2}x + 1)(-2x + 2) + g(x) \\ &= (\tfrac{1}{2}x + 1)\left[f(x) - (3x^2 - x - 1)g(x)\right] + g(x) \\ &= (\tfrac{1}{2}x + 1)f(x) - (\tfrac{3}{2}x^3 + \tfrac{5}{2}x^2 - \tfrac{3}{2}x - 2)g(x). \end{aligned}$$

Therefore, dividing by 3 we obtain $a(x) = \frac{1}{6}x + \frac{1}{3}$ and $b(x) = -\frac{1}{2}x^3 - \frac{5}{6}x^2 + \frac{1}{2}x + \frac{2}{3}$.

Once you have found polynomials $a(x)$ and $b(x)$, it is good practice to check they are correct by computing $af + bg$ directly.

Exercise 3.5

Let $f(x) = 2x^3 + 3x^2 + 2x - 1$ and $g(x) = x^2 + x$ be polynomials in $\mathbb{Q}[x]$.

(a) Find the highest common factor of f and g.

(b) Find polynomials $a(x)$ and $b(x)$ such that $af + bg = \text{hcf}(f, g)$.

3.4 Least common multiples

Throughout this section, we have been drawing parallels between the integers and polynomials with coefficients from fields. Here, our aim is to convert the results of Subsection 4.3 of Book A, Chapter 1.

Definition 3.10 *Least common multiple of two polynomials*

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then the **least common multiple** of f and g, $\operatorname{lcm}(f, g)$, is the monic polynomial $\ell \in F[x]$ satisfying the following:

(a) f and g both divide ℓ.

(b) For any polynomial $h \in F[x]$ that is divisible by both f and g, we have $\deg(h) \geq \deg(\ell)$.

First, we should establish whether $\operatorname{lcm}(f, g)$ must always exist. This is indeed the case because for any two non-zero polynomials f and g, the product fg and its associates satisfy condition (a) of Definition 3.10, and so there is a monic polynomial that satisfies condition (a). Among the monic polynomials satisfying (a), we need to choose one of smallest degree. Insisting that the least common multiple of two polynomials be monic then guarantees uniqueness, as we are about to show.

Lemma 3.11

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then $\ell = \operatorname{lcm}(f, g)$ is unique.

The proof follows the same pattern as the uniqueness part of the proof of Theorem 3.5.

Proof By the remarks before the lemma, it suffices to show that among the monic polynomials satisfying condition (a) of Definition 3.10, there is only one of lowest degree.

For a contradiction, suppose that ℓ_1 and ℓ_2 are two monic polynomials of the same lowest degree k that satisfy condition (a). However, then $\ell_1 - \ell_2$ is a polynomial of degree strictly less than k, and by Exercise 3.2(c), since each of f and g divides both ℓ_1 and ℓ_2, each also divides $\ell_1 - \ell_2$. Thus $\ell_1 - \ell_2$ is a polynomial of degree less than k that satisfies condition (a), which is a contradiction. ■

Now we have established that lcms are indeed unique, we want a practical way to find them. Of course, if we can factorise the polynomials f and g then it is relatively straightforward to find $\text{lcm}(f, g)$ (and indeed $\text{hcf}(f, g)$), simply by comparing factors as we would do with prime factorisations in the integers.

There are two problems with this approach. First, we haven't yet developed an analogue to 'prime factorisation' for polynomials (we will consider this task in the next section). Second, having managed to describe an analogue to 'prime factorisation', we would then need to find a way of calculating it. This is no easy task for integers, let alone polynomials over arbitrary fields! Instead, we will use the following result.

Proposition 3.12

Let $f(x), g(x)$ be non-zero monic polynomials with coefficients from a field F. Then

$$\text{lcm}(f, g) \cdot \text{hcf}(f, g) = f \cdot g.$$

Notice that this requires both f and g to be monic, otherwise fg is a scalar multiple of the left-hand side. Given non-monic polynomials f and g, we can obtain monic ones with the same lcm and hcf by multiplying each of f and g by the inverse of its leading coefficient.

Compare the proof of this proposition with the proof of Proposition 4.15 in Book A, Chapter 1: it is almost identical.

Proof Let $h(x) = \text{hcf}(f, g)$. Since h divides f, it also divides fg. Therefore, there exists a polynomial $\ell(x) \in F[x]$ such that $fg = h\ell$. We want to show that $\ell = \text{lcm}(f, g)$.

Since h divides both f and g, there exist polynomials p and q in $F[x]$ such that

$$f = hp \quad \text{and} \quad g = hq.$$

Remember that Lemma 2.10 of this chapter tells us that we are allowed to cancel factors from either side of an equation.

Substituting for f and then g in $fg = h\ell$, and then cancelling out a factor of h, we obtain

$$\ell = gp \quad \text{and} \quad \ell = fq.$$

This shows that ℓ is a common multiple of f and g, and so all that remains is for us to show it is the least such.

Suppose ℓ_1 is another polynomial in $F[x]$ that is divisible by both f and g. There exist polynomials $p_1, q_1 \in F[x]$ such that

$$\ell_1 = q_1 f \quad \text{and} \quad \ell_1 = p_1 g.$$

We also know from Theorem 3.5 that there exist polynomials a and b in $F[x]$ such that $h = \text{hcf}(f, g) = af + bg$.

Multiplying this equation by ℓ_1 gives

$$\begin{aligned}\ell_1 h &= \ell_1 af + \ell_1 bg \\ &= p_1 gaf + q_1 fbg \\ &= fg(p_1 a + q_1 b) \\ &= h\ell(p_1 a + q_1 b).\end{aligned}$$

This last line follows since we have already established that $fg = h\ell$. Finally, we can cancel a factor of h from either side to obtain

$$\ell_1 = \ell(p_1 a + q_1 b),$$

which proves that ℓ_1 is divisible by ℓ, and hence that $\ell(x) = \operatorname{lcm}(f, g)$. ■

Worked Exercise 3.13

Find $\operatorname{lcm}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2)$ in $\mathbb{Q}[x]$.

Solution

First, we compute $\operatorname{hcf}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2)$. The first round of polynomial long division tells us that

$$x^4 + 2x^3 - x - 2 = (x + 1)(x^3 + x^2 - x + 2) + (-2x - 4).$$

The remainder here is $-2x - 4 = -2(x + 2)$. Dividing $x^3 + x^2 - x + 2$ by the monic polynomial $x + 2$ gives

$$x^3 + x^2 - x + 2 = (x^2 - x + 1)(x + 2)$$

so that the hcf is $x + 2$.

Now, using Proposition 3.12,

$$\begin{aligned}&\operatorname{lcm}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2) \\ &\qquad = \frac{(x^3 + x^2 - x + 2) \cdot (x^4 + 2x^3 - x - 2)}{x + 2} \\ &\qquad = \frac{x^3 + x^2 - x + 2}{x + 2} \cdot (x^4 + 2x^3 - x - 2) \\ &\qquad = (x^2 - x + 1) \cdot (x^4 + 2x^3 - x - 2) \\ &\qquad = x^6 + x^5 - x^4 + x^3 - x^2 + x - 2.\end{aligned}$$

Exercise 3.6

Working in $\mathbb{Q}[x]$, find the lcm of:

(a) $x^3 - 1$ and $x^4 - x^3 + x^2 - 1$

(b) $x^3 + x^2 - 8x - 12$ and $x^3 + 5x^2 + 8x + 4$.

4 Factorising polynomials

This section explores methods of factorising polynomials into their 'indivisible' components.

4.1 Roots and irreducible polynomials

In the previous section, we learnt how to divide polynomials where the coefficients come from an arbitrary field, and this enabled us to compute hcfs and lcms. Another common task is to find roots: you will be familiar with exercises that ask for the values of $x \in \mathbb{R}$ for which $f(x) = 0$, and we will begin by generalising this to arbitrary fields.

Definition 4.1 *Root*

Let F be a field and let $f(x)$ be a non-zero polynomial in $F[x]$. Then $a \in F$ is called a **root** of $f(x)$ if $f(a) = 0$.

Notice that the above definition requires that the root a comes from the field F, that is, from the same field as the coefficients of the polynomial f. Why is this important? Consider, for example, the polynomial $f(x) = x^2 - 2$. As a polynomial in $\mathbb{R}[x]$, we can write $f(x) = (x - \sqrt{2})(x + \sqrt{2})$. On the other hand, as a polynomial in $\mathbb{Q}[x]$ we can see that $f(x)$ has no roots. We will explore this a little further in the next subsection, but note for now that the field over which a polynomial is defined (even if the coefficients come from a smaller field) really does matter.

Our first result should come as no surprise. It says that if we can find a root of a polynomial, then we can find a corresponding linear factor, and vice versa.

Theorem 4.2 *Factor Theorem*

Let F be a field, $a \in F$ and $f \in F[x]$ with $f \neq 0_x$. Then a is a root of f if, and only if, $f(x) = (x - a)q(x)$ for some $q(x) \in F[x]$, with $\deg(q) = \deg(f) - 1$.

Proof If $f(x) = (x - a)q(x)$, then clearly $f(a) = 0$.

For the converse, suppose that $f(a) = 0$. By the Division Algorithm (Theorem 3.1), there exist polynomials $q(x)$ and $r(x)$ such that $f(x) = (x - a)q(x) + r(x)$, where $\deg(r) < \deg(x - a) = 1$. Since $\deg(r) < 1$, $r(x)$ is a constant polynomial, and since $r(a) = 0$ it follows in fact that this constant must be zero, that is, $r(x) = 0_x$. Thus $f(x) = (x - a)q(x)$ as required. Finally, since $\deg(x - a) = 1$, we use Lemma 2.6 to obtain $\deg(f) = 1 + \deg(q)$. ■

We can deduce from the Factor Theorem that a polynomial in $F[x]$ of degree n has at most n roots. Loosely speaking, if a polynomial has more than n roots then it would have more than n factors, which if multiplied together would give a polynomial whose degree was greater than n.

Proposition 4.3

Let $f(x)$ be a non-zero polynomial of degree n in $F[x]$. Then the equation $f(x) = 0$ has at most n distinct roots in F.

Note that Lagrange's Theorem (Theorem 4.3 of Book A, Chapter 4) proves this result for the fields $\mathbb{Z}_p$, where p is prime.

Proof We will use induction on n, the degree of the polynomial. When $n = 1$, the polynomial is linear, and so we can write $f(x) = ax + b$. This has exactly one root as required.

Therefore, we will suppose $n \geq 2$, and suppose by induction that all polynomials of degree $m < n$ have at most m roots. If our polynomial $f(x)$ has no roots in F, then there is nothing more to prove, and so we may suppose that $a \in F$ is a root of $f(x)$. By the Factor Theorem 4.2, we have $f(x) = (x - a)g(x)$ for some polynomial $g(x) \in F[x]$ with $\deg(g) = n - 1$. By induction, $g(x)$ has at most $n - 1$ distinct roots in F.

Next, if $b \neq a$ is another root of $f(x)$, then it is also a root of $g(x)$. This follows because F has no zero divisors, and so, since $(b - a) \neq 0$, the equation $0 = f(b) = (b - a) \cdot g(b)$ implies that $g(b) = 0$. Thus, f cannot have more than n roots because otherwise g would have more than $n - 1$ roots, which is impossible. ■

Worked Exercise 4.4

Working in $\mathbb{Q}[x]$, factorise the following polynomials.

(a) $f(x) = x^3 + 3x^2 - 4$

(b) $f(x) = x^3 + 2x^2 + x$

Solution

(a) Note that $f(1) = 0$ and so $x - 1$ is a factor. Thus (using polynomial long division)

$$x^3 + 3x^2 - 4 = (x - 1)(x^2 + 4x + 4).$$

We can now factorise the remaining quadratic, either by inspection or using the quadratic equation, to get $f(x) = (x - 1)(x + 2)^2$.

(b) We can see immediately that x is a factor. Moreover, $f(-1) = 0$ and so $x + 1$ is a factor. We have $x^3 + 2x^2 + x = x(x + 1)(x + 1)$.

Exercise 4.1

Factorise the following polynomials over the stated field.

(a) $f(x) = x^4 - 3x^2 + 2$ over $\mathbb{R}$

(b) $f(x) = x^4 - 3x^2 + 2$ over $\mathbb{Q}$

(c) $f(x) = x^3 + x^2 + 2x + 2$ over $\mathbb{Z}_5$

(d) $f(x) = x^3 + x^2 + 2x + 2$ over $\mathbb{C}$

(e) $f(x) = x^4 - 6x^3 + 12x^2 - 10x + 3$ over $\mathbb{Q}$

Proposition 4.3 told us that a polynomial of degree n has at most n distinct roots, and we have already seen some examples where a polynomial has strictly fewer roots. As another example, consider the polynomial $x^2 + 1$: over $\mathbb{R}$ it has no roots, even though it is of degree 2. However, if we look at the same polynomial over $\mathbb{C}$ then we can find two roots, namely i and $-i$, and so

$$x^2 + 1 = (x + i)(x - i).$$

In fact, over the complex numbers a polynomial of degree n *always* has precisely n roots, provided we count repeated roots. This result is known as the Fundamental Theorem of Algebra and, ironically, is a theorem in complex analysis. We will not prove it here.

If we find, as we did in the case of $x^2 + 1$ over the reals, that for a given polynomial we do not identify as many roots as we might have expected, the question is whether we can nevertheless factorise the polynomial into smaller terms. For example, over $\mathbb{R}$ we can write the polynomial $x^4 - 1$ as a product

$$x^4 - 1 = (x + 1)(x - 1)(x^2 + 1),$$

and by what we saw a moment ago we know that $x^2 + 1$ does not factorise any further. In a sense, terms such as $x^2 + 1$ over $\mathbb{R}$ behave as the primes do for the integers, and this motivates our next definition.

Definition 4.5 *Irreducibility*

Let F be a field, and $f \in F[x]$ a non-zero polynomial over F. We say that f is **irreducible** if it is not a unit and it has no proper factors.

We say that f is **reducible** if it is not a unit but it does have proper factors.

For example, the polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$. However, since $x^2 + 1 = (x + i)(x - i)$, it is *not* irreducible in $\mathbb{C}[x]$. As we have said before, it is important to remember which polynomial ring we are working in!

In reading the above definition, we should recall *exactly* what we mean by 'unit' and 'proper factor' in a polynomial ring $F[x]$ over a field. First, the

units of $F[x]$ are, by Lemma 2.9 of this chapter, simply the non-zero elements of F. Next, if $f \neq 0_x$ can be written as $f = gh$ where g, h are 'proper factors', then neither g nor h is a unit; that is, the degrees of g and h must each be at least 1.

By virtue of the Fundamental Theorem of Algebra, the only irreducible polynomials in $\mathbb{C}[x]$ have degree 1. What about over $\mathbb{R}[x]$? The Fundamental Theorem of Algebra tells us something here too: suppose f is an irreducible polynomial in $\mathbb{R}[x]$ of degree greater than 1. As a polynomial in $\mathbb{C}[x]$, we can factorise it completely into linear factors. Moreover, if $z = a + ib$ $(b \neq 0)$ is a complex root of f, then the complex conjugate $\bar{z} = a - ib$ of z is also a non-real root of $f(x)$ in $\mathbb{C}$, since $f = \bar{f}$ is a polynomial with coefficients in $\mathbb{R}$. Thus, we can 'pair up' the complex linear factors of $f(x)$:

$$(x - (a + ib))(x - (a - ib)) = x^2 - 2ax + a^2 + b^2.$$

This product is of degree 2 and has coefficients from $\mathbb{R}$. Since it divides f and f is irreducible in $\mathbb{R}[x]$, it must be an associate of f. Thus $\deg(f) = 2$.

You might therefore be wondering where the interest in irreducible polynomials lies! The situation becomes much more interesting over the field of rationals $\mathbb{Q}$, and it is these that we will be exploring for the rest of this section.

Exercise 4.2

(a) Show that any polynomial of degree 1 is irreducible in $\mathbb{Q}[x]$.

(b) Prove that $f(x) = ax^2 + bx + c$ is irreducible in $\mathbb{R}[x]$ if, and only if, $b^2 - 4ac < 0$.

Recall that the expression $b^2 - 4ac$ is the discriminant of the quadratic polynomial $f(x)$.

4.2 Polynomials in $\mathbb{Z}[x]$ and rational roots

For the remainder of this section, we will focus on establishing conditions so that we can say when a polynomial in $\mathbb{Q}[x]$ is irreducible. You may have noticed that almost all the examples of polynomials we have seen in this chapter actually have integer coefficients; that is, they belong to the set

$$\mathbb{Z}[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{Z}, n \geq 0\}.$$

This is the **polynomial ring over** $\mathbb{Z}$. Checking that $\mathbb{Z}[x]$ is indeed a (commutative) ring follows the approach we used for $\mathbb{Q}[x]$, even though $\mathbb{Z}$ is not a field.

In addition to the fact that most of the polynomials we have encountered actually belong to $\mathbb{Z}[x]$, suppose that $f(x) = \sum_{i=0}^{n} a_ix^i$ is a polynomial in $\mathbb{Q}[x]$. This means that $a_i \in \mathbb{Q}$ for all $i = 0, 1, \ldots, n$, and so we can write $a_i = \dfrac{b_i}{c_i}$ where $b_i \in \mathbb{Z}$ and $c_i \in \mathbb{N}$ with $\text{hcf}(b_i, c_i) = 1$ for each i. If we multiply $f(x)$ by all of the denominators c_i, then we are left with a polynomial with integer coefficients. In other words, a polynomial in $\mathbb{Q}[x]$

has an associate that is a polynomial in $\mathbb{Z}[x]$. The following lemma ensures that the resulting polynomial has the same irreducibility properties as f.

Lemma 4.6

Let f, g be polynomials in $\mathbb{Q}[x]$, such that f and g are associates. Then f is irreducible if, and only if, g is irreducible.

We will prove this result via the contrapositive statement of the lemma: f is reducible if, and only if, g is reducible.

Proof Since f and g are associates, there exists $\frac{p}{q} \in \mathbb{Q}$ such that $g(x) = \frac{p}{q} f(x)$. Now, if $f(x)$ is reducible, then $f(x) = a(x)b(x)$ where a and b are polynomials in $\mathbb{Q}[x]$ each of degree at least 1. However, now we can write $g(x) = \left(\frac{p}{q} a(x)\right) \cdot b(x)$, which shows that $g(x)$ is also reducible. The reverse implication follows by a similar argument, after writing $f(x) = \frac{q}{p} g(x)$. ■

Therefore, we can now restrict our attention to polynomials with integer coefficients, while continuing to consider irreducibility over $\mathbb{Q}[x]$. We will begin with a simple test that will enable us to determine whether a polynomial in $\mathbb{Z}[x]$ has any roots in $\mathbb{Q}[x]$.

Theorem 4.7 *Rational Root Test*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$ of degree $n \geq 1$, with a_n and a_0 non-zero. Then any rational root $\frac{p}{q} \in \mathbb{Q}$ of $f(x)$ with $\text{hcf}(p, q) = 1$ must satisfy the following:

(a) p is an integer factor of a_0

(b) q is an integer factor of a_n.

Proof Since $f\left(\frac{p}{q}\right) = 0$, we have

$$a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + \cdots + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Moving the constant term a_0 to the right-hand side, and multiplying the equation by q^n, we obtain

$$p\left(a_n p^{n-1} + a_{n-1} q p^{n-2} + \cdots + a_1 q^{n-1}\right) = -a_0 q^n.$$

On the left-hand side, the expression in brackets is an integer (since all the terms are integers) and therefore the left-hand side is a multiple of p. This implies that p must divide $a_0 q^n$, but since $\text{hcf}(p, q) = 1$, it follows that p divides a_0.

Similarly, instead of moving the constant term to the right-hand side, we can move the leading term over, to obtain

$$q\left(a_{n-1} p^{n-1} + a_{n-2} q p^{n-2} + \cdots + a_0 q^{n-1}\right) = -a_n p^n.$$

By the same argument as before, q must divide $a_n p^n$, but since $\text{hcf}(p, q) = 1$ we must have $q \mid a_n$. ■

This theorem provides us with a method to find all rational roots of a polynomial in $\mathbb{Z}[x]$, and therefore of any polynomial in $\mathbb{Q}[x]$.

Worked Exercise 4.8

Find all the rational roots of $f(x) = 2x^3 + 3x - 5$.

Solution

By Theorem 4.7, if $\frac{p}{q} \in \mathbb{Q}$ is a solution to $f(x) = 0$, then p must divide 5 and q must divide 2.

Therefore, the possible values for p are ± 1, ± 5, and the possible values for q are 1, 2. This implies that the list of values of $\frac{p}{q}$ that *could* be roots of $f(x)$ is as follows.

Notice that we need to include negative values of p too.

$$\begin{aligned}\frac{p}{q} &= \pm\tfrac{1}{1}, \pm\tfrac{1}{2}, \pm\tfrac{5}{1}, \pm\tfrac{5}{2} \\ &= \pm 1, \pm\tfrac{1}{2}, \pm 5, \pm\tfrac{5}{2}\end{aligned}$$

We now need to check each possibility in turn, to determine whether any of these values are roots:

$$\begin{array}{llll} f(1) = 0 & f(-1) = -10 & f(\tfrac{1}{2}) = -\tfrac{13}{4} & f(-\tfrac{1}{2}) = -\tfrac{27}{4} \\ f(5) = 260 & f(-5) = -270 & f(\tfrac{5}{2}) = \tfrac{135}{4} & f(-\tfrac{5}{2}) = -\tfrac{175}{4}. \end{array}$$

From this, we conclude that the only root of $f(x)$ in $\mathbb{Q}$ is $x = 1$.

Exercise 4.3

Find the roots, if any, in $\mathbb{Q}$ of:

(a) $4x^4 - 2x^3 - 2x^2 - x + 1$

(b) $x^4 - x^3 - 2x^2 - 2x + 4$.

Exercise 4.4

Is $f(x) = 2x^3 - x^2 + 4x + 1$ irreducible in $\mathbb{Q}[x]$?

4.3 Gauss's Lemma

Given a polynomial in $\mathbb{Z}[x]$, can we determine whether it is irreducible in $\mathbb{Q}[x]$ by looking at its irreducibility in $\mathbb{Z}[x]$? The main result of this subsection, Gauss's Lemma, will tell us that we can. However, first we need to say what we mean by 'irreducible' in $\mathbb{Z}[x]$.

Definition 4.9 *Irreducibility over $\mathbb{Z}$*

Let f, g be polynomials in $\mathbb{Z}[x]$, with $f \neq 0_x$.

(a) We say that g **divides** f (or is a **factor** of f) if there exists some $h \in \mathbb{Z}[x]$ such that $f = gh$.

(b) If $f = gh$ and neither g nor h is equal to ± 1, then g, h are **proper factors**.

(c) We say that f is **irreducible** over $\mathbb{Z}$ if $f \neq \pm 1$ and it has no proper factors in $\mathbb{Z}[x]$.

Recall that the units of $\mathbb{Z}$ are 1 and -1.

We have changed very little to extend our definitions to polynomials over $\mathbb{Z}$, but there is one important point that you need to be aware of.

Example 4.10 *Factorisation in different polynomial rings*

Consider the polynomial $f(x) = 2x^2 + 2$.

(a) In $\mathbb{Z}[x]$, the factorisation $f(x) = 2 \cdot (x^2 + 1)$ is a proper factorisation, since neither factor is equal to ± 1, which are the units of $\mathbb{Z}$.

(b) In $\mathbb{Q}[x]$, the factorisation $f(x) = 2 \cdot (x^2 + 1)$ is *not* proper, since 2 is a unit. In fact, $f(x)$ is an associate of $x^2 + 1$, which is irreducible in $\mathbb{Q}[x]$ by Lemma 4.6.

At first sight, the above example might suggest that irreducibility in $\mathbb{Q}[x]$ is very different from irreducibility in $\mathbb{Z}[x]$. However, the following two definitions will enable us to proceed.

Definition 4.11 *Content of a polynomial*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$.

(a) The **content** of f is the highest common factor of all its coefficients:

$$c_f = \text{hcf}(a_0, a_1, \ldots, a_n).$$

(b) $f(x)$ is **primitive** if $c_f = 1$, or, in other words, if its coefficients have no non-trivial common factor.

For example, the content of the polynomial $6x^3 + 9x^2 - 3x + 18$ is $\text{hcf}(6, 9, -3, 18) = 3$.

We can eliminate the discrepancy encountered in Example 4.10 by restricting our view to primitive polynomials in $\mathbb{Z}[x]$. Note that if f is a polynomial in $\mathbb{Z}[x]$, then $f(x) = c_f f^*(x)$ for some primitive polynomial $f^*(x) \in \mathbb{Z}[x]$. Viewing f and f^* as polynomials over $\mathbb{Q}$, we can see that they are associates since the content c_f is an integer, and therefore a rational number. By Lemma 4.6, f^* is irreducible in $\mathbb{Q}[x]$ if, and only if, f is. So, from now on it suffices to prove results relating to irreducibility

using primitive polynomials only, and then the extension to all polynomials follows directly.

We require one last result before we state and prove Gauss's Lemma. The following statement looks innocent enough, but you may be surprised by how much work is required to prove it!

Lemma 4.12

The polynomials f and g are primitive in $\mathbb{Z}[x]$ if, and only if, their product fg is a primitive polynomial in $\mathbb{Z}[x]$.

Proof First we prove that if fg is primitive then f and g must also be primitive. Heading for a contradiction, suppose (without loss of generality) that f is not primitive. Then by writing $f = c_f f^*$, where c_f is the (non-unit) content of f and f^* is a primitive polynomial in $\mathbb{Z}[x]$, we see that $fg = (c_f f^*)g = c_f(f^*g)$ is also not primitive, proving one direction of the lemma.

Conversely, suppose that $f(x) = \sum_{i=0}^{n} a_i x^i$ and $g(x) = \sum_{i=0}^{m} b_i x^i$ are two primitive polynomials in $\mathbb{Z}[x]$, but that their product fg is not primitive. Therefore there exists a prime number $p \in \mathbb{Z}$ that divides the content of fg.

Since f and g are primitive, we know that p does not divide all the coefficients of f or of g. Therefore, we can find coefficients a_r of f and b_s of g such that p does not divide a_r or b_s. We pick r and s as large as possible, so that every coefficient of a higher-power term in each of f and g is divisible by p. Now consider the coefficient of x^{r+s} in fg:

$$\sum_{i+j=r+s} a_i b_j = a_0 b_{r+s} + a_1 b_{r+s-1} + \cdots + a_r b_s + \cdots + a_{r+s} b_0.$$

In this sum, when $i = r$ and $j = s$ then $a_i b_j$ is not divisible by p. However, for every other term $a_i b_j$ we have either $i > r$ or $j > s$, and so it must be divisible by p. Consequently, the coefficient of x^{r+s} cannot be divisible by p, which contradicts the fact that p divides the content of fg. Therefore, fg must be primitive. ■

Consider a primitive polynomial $f \in \mathbb{Z}[x]$: if $f = gh$, where $g, h \in \mathbb{Z}[x]$ are proper factors of f, then g and h are also proper factors of f in $\mathbb{Q}[x]$. Therefore, if f is reducible in $\mathbb{Z}[x]$, it is also reducible in $\mathbb{Q}[x]$. Gauss's Lemma proves the converse to this.

Theorem 4.13 *Gauss's Lemma*

Let $f(x)$ be a primitive polynomial in $\mathbb{Z}[x]$. If we can write $f = gh$ with $g, h \in \mathbb{Q}[x]$, then we can write $f = GH$ with $G, H \in \mathbb{Z}[x]$, $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Importantly, note that $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$, so that the proper factorisations of f in $\mathbb{Q}[x]$ and $\mathbb{Z}[x]$ are going to be closely connected.

Proof Since $g(x)$ is a polynomial in $\mathbb{Q}[x]$ we can find the lowest common denominator, b say, of all its coefficients. Hence we can write $g(x) = \frac{a}{b}G(x)$, where $G(x)$ is a primitive polynomial in $\mathbb{Z}[x]$ and $\frac{a}{b}$ is a fraction in its lowest terms. Similarly we can write $h(x) = \frac{c}{d}H(x)$, where $H(x)$ is a primitive polynomial in $\mathbb{Z}[x]$ and $\frac{c}{d}$ is a fraction in its lowest terms. Note that $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Hence $f(x) = g(x)h(x) = \frac{ac}{bd}G(x)H(x)$: we want to prove that $f = GH$, and so we need to show that $ac = bd$. Rearranging, we find $bdf = acGH$. Since f is primitive, the content of $bdf(x)$ is bd. Moreover, by Lemma 4.12 GH is primitive, and so the content of $acG(x)H(x)$ is ac. Since the polynomials $bdf(x)$ and $acG(x)H(x)$ are equal, so must their contents be equal, and hence $ac = bd$, as required. ■

Worked Exercise 4.14

Show that $f(x) = x^4 + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Solution

Note that if $f(x)$ has a factor of degree 3, it also has a factor of degree 1.

We need to show that $f(x)$ has no factors of degrees 1 or 2.

First we will eliminate the possibility of linear factors, using the Factor Theorem 4.2 and the Rational Root Test, Theorem 4.7. The latter tells us that any root $\frac{p}{q} \in \mathbb{Q}$ of $f(x)$ must have $p \mid 1$ and $q \mid 1$. Thus $p = \pm 1$, $q = 1$, and so the only numbers we need to check are ± 1. It is easy to verify that neither is a root of $f(x)$, and hence $f(x)$ has no linear factors.

This leaves the possibility that $f(x)$ factorises into two quadratic terms. By Gauss's Lemma, if $f(x)$ is reducible in $\mathbb{Q}[x]$ then it is reducible in $\mathbb{Z}[x]$, and so we can assume

$$f(x) = (ax^2 + bx + c)(rx^2 + sx + t)$$

where $a, b, c, r, s, t \in \mathbb{Z}$.

Multiplying out, we have

$$x^4 + x + 1 = arx^4 + (as + br)x^3 + (at + bs + cr)x^2 + (bt + cs)x + ct.$$

We can now compare coefficients between the left-hand side and the right-hand side. First, the coefficient of x^4 tells us that $ar = 1$, which implies that $a = r = \pm 1$. By multiplying all of a, b, c, r, s, t by -1 if necessary, we can assume $a = r = 1$.

Next, we consider the constant term, which tells us that $ct = 1$, and so $c = t = \pm 1$. Substituting $a = r = 1$ and $c = t = \pm 1$ into the equation above gives

$$x^4 + x + 1 = x^4 + (s + b)x^3 + (\pm 1 + bs \pm 1)x^2 \pm (b + s)x + 1.$$

The coefficient of x^2 gives $0 = \pm 2 + bs$, and the coefficient of x^3 gives $0 = s + b$, and so $b = -s$. Eliminating s, we find $b^2 = \pm 2$.

Neither $b^2 = 2$ nor $b^2 = -2$ has a solution in $\mathbb{Z}$, and so there can be no factorisation of $f(x)$ into quadratics in $\mathbb{Z}[x]$. Therefore $f(x)$ is irreducible in $\mathbb{Z}[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Gauss's Lemma.

Exercise 4.5

Is $f(x) = x^4 - x^2 + 1$ irreducible in $\mathbb{Q}[x]$?

4.4 Further tests for irreducibility

Consider the polynomial $2x^5 - 3x^3 + 9x + 6$. Is it irreducible in $\mathbb{Q}[x]$? To answer this, we could modify the technique used in Worked Exercise 4.14, but the next theorem gives us a much quicker method.

Theorem 4.15 *Eisenstein's Criterion*

Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $\mathbb{Z}[x]$. If there is a prime $p \in \mathbb{Z}$ such that

(a) p divides a_i for $i = 0, \ldots, n-1$, but p does not divide a_n, and

(b) p^2 does not divide a_0,

then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof First, we may assume that $f(x)$ is primitive in $\mathbb{Z}[x]$ by the comments after Definition 4.11. By Gauss's Lemma, if $f = gh$ in $\mathbb{Q}[x]$ then $f = GH$ for some $G, H \in \mathbb{Z}[x]$ such that $\deg(G) = \deg(g)$ and $\deg(H) = \deg(h)$. Suppose that $G(x) = \sum_{i=0}^{r} b_i x^i$ and $H(x) = \sum_{j=0}^{s} c_j x^j$. Then, comparing coefficients gives

$$a_0 = b_0 c_0$$
$$a_1 = b_0 c_1 + b_1 c_0$$
$$a_k = \sum_{i=0}^{k} b_i c_{k-i} = b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1 + b_k c_0,$$
for each $k \in \{0, \ldots, n\}$.

(We set $b_i = 0$ for $i = r+1, \ldots, n$ and $c_j = 0$ for $j = s+1, \ldots, n$.)

Since p divides a_0, p divides either b_0 or c_0. Since p^2 does not divide a_0, p will not divide both b_0 and c_0. Suppose, without loss of generality, that p divides b_0 and not c_0.

Next, since p divides a_1 and not c_0, the second equation above shows that p divides b_1.

Now suppose that b_k is the first coefficient in $G(x)$ that is not divisible by p. Since

$$a_k = b_0c_k + b_1c_{k-1} + \cdots + b_{k-1}c_1 + b_kc_0,$$

only the last term in this sum is not divisible by p, and therefore p does not divide a_k. This is possible only if $k = n$, which implies that $b_n \neq 0$ and therefore $\deg(G) = n$. It follows that $\deg(g) = n$ and so $f(x)$ is irreducible in $\mathbb{Q}[x]$. ■

Worked Exercise 4.16

Show that the polynomial $2x^5 - 3x^3 + 9x + 6$ is irreducible in $\mathbb{Q}[x]$.

Solution

We apply Eisenstein's Criterion: $p = 3$ is a prime number and divides all the coefficients except the leading coefficient. $3^2 = 9$ does not divide 6, the constant term. Thus the conditions of Eisenstein's Criterion are satisfied and the polynomial is irreducible in $\mathbb{Q}[x]$.

Exercise 4.6

(a) Show that $x^n - p$ is irreducible in $\mathbb{Q}[x]$ for any prime number p.

(b) Show that $f(x) = \frac{3}{5}x^5 + \frac{5}{3}x^4 - x^3 + \frac{10}{3}x^2 + 2x - \frac{1}{3}$ is irreducible in $\mathbb{Q}[x]$.

One important family of polynomials in number theory is the **cyclotomic polynomials**, defined by

$$\Phi_n(x) = x^{n-1} + x^{n-2} + \cdots + x + 1 = \frac{x^n - 1}{x - 1}.$$

Φ is the Greek letter 'capital phi'.

Notice that the roots of $\Phi_n(x)$ in $\mathbb{C}$ satisfy $x^n = 1$ but $x \neq 1$ and thus consist of all the nth roots of unity, except for 1 itself. We will be interested in the case where $n = p$, a prime number, and we will use Eisenstein's Criterion to show that $\Phi_p(x)$ is irreducible in $\mathbb{Q}[x]$. However, we cannot use Eisenstein directly because all the coefficients of $\Phi_p(x)$ are equal to 1. Instead, we need a trick.

Lemma 4.17

Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$ if, and only if, $f(x-b)$ is irreducible in $\mathbb{Q}[x]$ for any $b \in \mathbb{Q}$.

Proof Suppose $f(x) = g(x)h(x)$ in $\mathbb{Q}[x]$, where $g(x)$ and $h(x)$ are proper factors. Then $f(x-b) = g(x-b)h(x-b)$ is a proper factorisation of $f(x-b)$, since $\deg(g(x-b)) = \deg(g(x))$ and $\deg(h(x-b)) = \deg(h(x))$.

The reverse direction follows similarly: if $f(x-b)$ is reducible, then $f(x-b) = g(x-b)h(x-b)$ for some proper factors g, h. Setting $y = x-b$ we obtain $f(y) = g(y)h(y)$. From this we can deduce that $f(y-(-b)) = g(y-(-b))h(y-(-b))$. But $y-(-b) = x$ and so $f(x) = g(x)h(x)$. ■

Worked Exercise 4.18

Let p be a prime number. Prove that the cyclotomic polynomial $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Solution

First, recall that $\Phi_p(x) = \dfrac{x^p - 1}{x - 1}$ and therefore, by applying the binomial theorem,

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{x^p + \binom{p}{1}x^{p-1} + \binom{p}{2}x^{p-2} + \cdots + \binom{p}{p-1}x}{x}$$

$$= x^{p-1} + \binom{p}{1}x^{p-2} + \binom{p}{2}x^{p-3} + \cdots + \binom{p}{p-1}.$$

Now $p \mid \binom{p}{i}$ for $1 \leq i \leq p-1$, and p^2 does not divide $\binom{p}{p-1} = p$, and so $\Phi_p(x+1)$ satisfies Eisenstein's Criterion. Therefore $\Phi_p(x+1)$ is irreducible in $\mathbb{Q}[x]$, and thus by Lemma 4.17 so also is $\Phi_p(x)$.

The **binomial coefficient** is defined as $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$.

Exercise 4.7

Prove that the following polynomials are irreducible in $\mathbb{Q}[x]$.

(a) $x^4 + 1$

(b) $x^8 + 1$

We finish with one final method for determining irreducibility of polynomials in $\mathbb{Z}[x]$, and hence, by Gauss's Lemma, in $\mathbb{Q}[x]$. This technique involves reducing the coefficients of the polynomial modulo some prime p.

Lemma 4.19

Let f be a polynomial in $\mathbb{Z}[x]$ and let p be a prime that does not divide the leading coefficient of f. If f is irreducible in $\mathbb{Z}_p[x]$, then f is irreducible in $\mathbb{Z}[x]$.

Before we prove this, consider the following example, which proves that the reverse of this lemma is not true.

Example 4.20 *Irreducibility in different polynomial rings*

The polynomial $x^2 + 3$ is irreducible in $\mathbb{Z}[x]$, but in $\mathbb{Z}_2[x]$ we have

$$x^2 + 3 \equiv x^2 + 1 \equiv x^2 + 2x + 1 \equiv (x + 1)^2 \pmod 2.$$

So $x^2 + 3$ is reducible in $\mathbb{Z}_2[x]$.

Note, however, that $x^2 + 3$ is irreducible in $\mathbb{Z}_5[x]$ since any proper factor must be a linear factor, and the equation $x^2 \equiv -3 \equiv 2 \pmod 5$ has no solution since 2 is not a quadratic residue modulo 5.

Proof of Lemma 4.19 Working towards a contradiction, suppose that f is irreducible in $\mathbb{Z}_p[x]$, but that $f = gh$ is a proper factorisation of f in $\mathbb{Z}[x]$. Let g^* and h^* denote the polynomials obtained from g and h by reducing the coefficients modulo p. We then have $f(x) \equiv g^*(x)h^*(x) \pmod p$, and since the leading coefficient of f is not divisible by p, it follows that $\deg(g) = \deg(g^*)$ and $\deg(h) = \deg(h^*)$, and therefore f is reducible in $\mathbb{Z}_p[x]$, which is a contradiction. ■

To use this lemma, the skill is to find a prime p such that reducing modulo p gives a polynomial that can be shown to be irreducible in $\mathbb{Z}_p[x]$.

Worked Exercise 4.21

Show that $f(x) = x^4 + 15x^2 + 7$ is irreducible in $\mathbb{Q}[x]$.

Solution

Note first that if we reduce modulo 2, we get

$$f(x) \equiv x^4 + x^2 + 1 \equiv (x^2 + x + 1)(x^2 + x + 1) \pmod 2,$$

which is not irreducible. Similarly, reducing modulo 3 also does not work because

$$f(x) \equiv x^4 + 1 \equiv (x^2 + x - 1)(x^2 - x - 1) \pmod 3.$$

Next, we try reducing modulo 5 to get

$$f(x) \equiv x^4 + 2 \pmod 5,$$

which is not obviously reducible. First, we test whether any of the five elements of $\mathbb{Z}_5$ is a root of $f(x)$ in $\mathbb{Z}_5[x]$, and we find that none is:

$$f(0) \equiv 2, \quad f(1) \equiv 3, \quad f(2) \equiv 3, \quad f(3) \equiv 3, \quad f(4) \equiv 3.$$

This leaves the possibility that $f(x) \equiv g(x)h(x) \pmod 5$, where g and h are quadratic factors in $\mathbb{Z}_5[x]$. Since f is monic, we can assume without loss of generality that both g and h are monic. So then we can write

$$x^4 + 2 \equiv (x^2 + ax + b)(x^2 + cx + d)$$

for some $a, b, c, d \in \mathbb{Z}_5$. Comparing coefficients of x^3, x^2 and x^0 in turn, we get the following equations:

$$\begin{aligned} a + c &\equiv 0 \pmod 5 \\ ac + b + d &\equiv 0 \pmod 5 \\ bd &\equiv 2 \pmod 5. \end{aligned}$$

Using $c \equiv -a$ from the first equation in the second tells us that $b + d \equiv a^2$, and so $b + d \equiv 0, 1$ or 4 since these are the only quadratic residues modulo 5. It follows that $d \equiv -b, 1 - b$ or $4 - b$. Using these three possible values of d in the third equation, $bd \equiv 2$, gives

$$\begin{aligned} b^2 &\equiv -2 \pmod 5 \\ b(1 - b) &\equiv 2 \pmod 5 \\ b(4 - b) &\equiv 2 \pmod 5. \end{aligned}$$

We can exhaustively check that none of the values $b = 0, 1, 2, 3, 4$ satisfies any of these equations. This tells us that $x^4 + 2$ has no quadratic factor modulo 5, and therefore that $f(x)$ is irreducible modulo 5. Hence, by Lemma 4.19, $f(x)$ is irreducible in $\mathbb{Z}[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Gauss's Lemma.

Exercise 4.8

Are the following polynomials irreducible in $\mathbb{Q}[x]$? If not, give their factorisations into irreducible polynomials.

(a) $x^3 + 7x^2 - 14x + 4$

(b) $x^4 + 4$

Exercise 4.9

(a) List the irreducible polynomials of degree 4 or less in $\mathbb{Z}_2[x]$.

(b) Hence, or otherwise, prove that the following polynomials are irreducible over $\mathbb{Q}$.

(i) $x^4 - 3x^3 + 12x^2 - 2x + 5$

(ii) $3x^4 - x^3 + 71x^2 - 9x + 1$

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should:

(a) Recall and be able to use the axioms that define a ring, and know the basic properties of rings arising from these axioms.

(b) Understand the definition of 'subring', and the statement, proof and uses of the subring criterion.

(c) Know the meaning of and be able to use the terms 'zero divisor' and 'unit'.

(d) Understand the terms 'field' and 'subfield', and recall basic properties of these.

(e) Know how to add and multiply polynomials over arbitrary fields, and be able to use this to define polynomial rings.

(f) Understand the term 'degree of a polynomial', and recall results relating the degree of polynomials to the addition and multiplication operations.

(g) Recall basic properties of polynomial rings, such as the non-existence of zero divisors, classification of units and cancellation of polynomials.

(h) Understand the terms 'associate' and 'monic polynomial'.

(i) Understand the statement and proof of the Division Algorithm for polynomials, and be able to apply polynomial long division in the ring $\mathbb{Q}[x]$.

(j) Understand the meaning of the highest common factor of two polynomials, the proof of the existence of the hcf, and the meaning of 'coprime' in the context of polynomials over fields.

(k) Be able to apply the Euclidean Algorithm to compute the hcf of two polynomials f and g in $\mathbb{Q}[x]$, and find polynomials a, b such that $\mathrm{hcf}(f, g) = af + bg$.

(l) Understand the meaning of the least common multiple of two polynomials, the proof of its uniqueness, and be able to compute lcms in the polynomial ring $\mathbb{Q}[x]$.

(m) Understand the terms 'root', 'irreducible' and 'reducible' polynomial, and be able to recall and use the Factor Theorem to factorise polynomials defined over arbitrary fields.

(n) Understand and be able to apply the Rational Root Test to find roots of polynomials over $\mathbb{Q}$.

(o) Understand the terms 'content' and 'primitive', and recall the statement and proof of Gauss's Lemma.

(p) Be able to apply Gauss's Lemma, Eisenstein's Criterion and its variants to test for irreducibility of polynomials in $\mathbb{Q}[x]$.

Solutions to exercises

Solution to Exercise 1.1

(a) $(2,1)+(4,5)=(6,6)$ and $(2,1)\cdot(4,5)=(8,5)$

(b) $(0,0)$

(c) $(-2,-1)$

(d) $(1,1)$

Solution to Exercise 1.2

For the additive axioms, we could use the fact that $(\mathbb{Z}_n,+)$ is an abelian group.

(a) We check each axiom in turn. R1, R2 and R5 follow by the basic properties of addition modulo n. For R3, the additive identity is 0, since $0+a\equiv a\pmod{n}$ for any $a\in\mathbb{Z}_n$, and then writing $a+(-a)\equiv 0\pmod{n}$ proves R4.

The multiplicative axioms follow similarly: R6 and R7 follow directly from the definition of multiplication modulo n. For R8, the multiplicative identity is 1, since $1\cdot a=a\cdot 1\equiv a\pmod{n}$.

Finally, R9 follows directly from the distributivity of addition and multiplication of the integers, and R10 (commutativity of multiplication) holds by the properties of multiplication modulo n. Hence $(\mathbb{Z}_n,+,\cdot)$ is a commutative ring.

(b) For $\mathbb{Z}[\sqrt{2}]$, let $a,b,c,d\in\mathbb{Z}$. We consider each axiom in turn.

For R1, we have

$$(a+b\sqrt{2})+(c+d\sqrt{2})=(a+c)+(b+d)\sqrt{2},$$

which is in $\mathbb{Z}(\sqrt{2})$ since $a+c,b+d\in\mathbb{Z}$.

Using the equation in R1 repeatedly together with the associativity of addition of integers, R2 follows directly. The additive identity is $0=0+0\sqrt{2}$, which proves R3, and then for R4, the inverse of $a+b\sqrt{2}$ is $(-a)+(-b)\sqrt{2}$. Commutativity follows from the equation in R1 with the commutativity of addition over the real numbers.

For the multiplicative axioms, closure (R6) is not immediately obvious, but it follows from

$$\begin{aligned}(a+b\sqrt{2})\cdot(c+d\sqrt{2}) &= ac+(bc+ad)\sqrt{2}+bd\sqrt{2}\cdot\sqrt{2}\\ &= (ac+2bd)+(bc+ad)\sqrt{2}.\end{aligned}$$

R7 and R10 are now immediate since multiplication of real numbers is associative and commutative. For R8, the multiplicative identity is $1=1+0\sqrt{2}$. Finally, R9 follows, once again, because multiplication is distributive over addition in $\mathbb{R}$.

Solution to Exercise 1.3

We use part (c) of the proposition twice, followed by part (b):

$$(-a)\cdot(-b)=-(a\cdot(-b))=-(-(a\cdot b))=a\cdot b.$$

Solution to Exercise 1.4

If $1 = 0$ and a is any element of the ring then, using Proposition 1.5(a),

$$a = a \cdot 1 = a \cdot 0 = 0$$

and hence $a = 0$. Consequently, the ring has only one element, which must be both the additive and the multiplicative identity.

Solution to Exercise 1.5

(a) We show that $\mathbb{Z}[\omega]$ is a subring of $\mathbb{C}$. Let $a, b, c, d, e, f \in \mathbb{Z}$. For SR1, we have

$$(a + b\omega + c\omega^2) - (d + e\omega + f\omega^2) = (a - d) + (b - e)\omega + (c - f)\omega^2,$$

which is in $\mathbb{Z}[\omega]$ since $(a - d)$, $(b - e)$ and $(c - f) \in \mathbb{Z}$.

For SR2, using $\omega^3 = 1$, we have

$$\begin{aligned}&(a + b\omega + c\omega^2) \cdot (d + e\omega + f\omega^2)\\ &\quad = (ad + bf + ce) + (ae + bd + cf)\omega + (af + be + cd)\omega^2 \in \mathbb{Z}[\omega].\end{aligned}$$

Finally, the multiplicative identity of $\mathbb{C}$ is 1, and in $\mathbb{Z}[\omega]$ we have $1 = 1 + 0\omega + 0\omega^2$. Hence SR3 holds.

Note that $\mathbb{Z}[\omega]$ is *not* a subring of $\mathbb{Z}[i]$.

(b) We will show that $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. For a, b, c, d in $\mathbb{Q}$, we have

SR1: $(a + ib) - (c + id) = (a - c) + i(b - d) \in \mathbb{Q}[i]$.

SR2: $(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc) \in \mathbb{Q}[i]$.

SR3: The multiplicative identity of $\mathbb{C}$ and $\mathbb{Q}[i]$ is $1 = 1 + 0i$.

Solution to Exercise 1.6

(a) $(3, 0) \cdot (0, 1) = (0, 0)$ gives an example of two zero divisors.

(b) The zero divisors are those elements (x, y) for which one of x and y is 0.

Solution to Exercise 1.7

(a) The units are 1, 5, 7 and 11. Each element is its own inverse – for example, $7 \times 7 \equiv 1 \pmod{12}$.

(b) The units are the pairs $(1, 1)$, $(1, -1)$, $(-1, 1)$ and $(-1, -1)$.

(c) The units are $1, -1, i$ and $-i$.

Solution to Exercise 1.8

(a) The zero divisors in $\mathbb{Z}_8$ are 2, 4 and 6, since $2 \times 4 \equiv 6 \times 4 \equiv 0 \,(\text{mod } 8)$.

In $\mathbb{Z}_9$ the zero divisors are 3 and 6.

In $\mathbb{Z}_{10}$ the zero divisors are 2, 4, 5, 6, 8.

In $\mathbb{Z}_{11}$ there are no zero divisors.

In $\mathbb{Z}_{12}$ the zero divisors are 2, 3, 4, 6, 8, 9 and 10.

We will use this result in the proof of Lemma 1.18.

(b) The zero divisors in $\mathbb{Z}_n$ are those positive integers $a \in \mathbb{Z}_n$ that share a common factor with n that is greater than 1; that is, $\text{hcf}(a, n) > 1$.

(c) $\mathbb{Z}_n$ has no zero divisors if, and only if, n is a prime number.

(d) Suppose that $a \in R$ is a unit, and so there exists a multiplicative inverse a^{-1}. If there exists $b \in R$ such that $a \cdot b = 0$, then

$$b = 1 \cdot b = (a^{-1}a)b = a^{-1}(ab) = a^{-1} \cdot 0 = 0.$$

Therefore a cannot be a zero divisor.

Solution to Exercise 1.9

(a) By definition, $(a^{-1})^{-1} \cdot a^{-1} = 1$, and therefore we have

$$\begin{aligned}(a^{-1})^{-1} &= (a^{-1})^{-1} \cdot 1 \\ &= (a^{-1})^{-1} \cdot (a^{-1} \cdot a) \\ &= ((a^{-1})^{-1} \cdot a^{-1}) \cdot a \\ &= 1 \cdot a = a.\end{aligned}$$

(b)
$$\begin{aligned}b^{-1}a^{-1} &= (b^{-1}a^{-1}) \cdot 1 = (b^{-1}a^{-1}) \cdot \big((ab)(ab)^{-1}\big) \\ &= \big((b^{-1}a^{-1})(ab)\big) \cdot (ab)^{-1} \\ &= \big(b^{-1}(a^{-1}a)b\big) \cdot (ab)^{-1} = \big(b^{-1}b\big) \cdot (ab)^{-1} \\ &= 1 \cdot (ab)^{-1} = (ab)^{-1}.\end{aligned}$$

Alternatively, consider

$$\begin{aligned}(b^{-1}a^{-1}) \cdot (ab) &= b^{-1}(a^{-1}a)b \\ &= b^{-1} \cdot 1 \cdot b = b^{-1}b \\ &= 1,\end{aligned}$$

which shows that $(b^{-1}a^{-1})$ is the inverse of the element ab. That is, $(ab)^{-1} = b^{-1}a^{-1}$.

Solution to Exercise 1.10

(a) From Exercise 1.5(b), we know that $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. To show that $\mathbb{Q}[i]$ is a subfield of $\mathbb{C}$, it remains to show only that every non-zero element of $\mathbb{Q}[i]$ has a multiplicative inverse, by Lemma 1.21. Let $a, b \in \mathbb{Q}$, not both equal to zero, so that $a + bi \in \mathbb{Q}[i]$ is non-zero.

We want to find $c, d \in \mathbb{Q}$ such that $(a+bi)(c+di) = 1$. We could compare real and imaginary parts, and solve the resulting simultaneous equations, but there is a quicker method. Observe that $(a+bi)(a-bi) = a^2 + b^2$ is a rational number, and therefore

$$\frac{(a+bi)(a-bi)}{a^2+b^2} = 1.$$

Thus $c = \dfrac{a}{a^2+b^2}$ and $d = \dfrac{-b}{a^2+b^2}$ are both in $\mathbb{Q}$, and the multiplicative inverse $c + di \in \mathbb{Q}[i]$ is

$$\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

(b) We will use the fact that $\mathbb{Q}[\sqrt{2}] \subset \mathbb{R}$. First we show, using the subring criterion (Lemma 1.8) that $\mathbb{Q}[\sqrt{2}]$ is a subring of $\mathbb{R}$. For $a, b, c, d \in \mathbb{Q}$, we have

SR1: $(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a-c) + (b-d)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$.

SR2: $(a + b\sqrt{2}) \cdot (c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$.

SR3: The multiplicative identity of $\mathbb{R}$ and $\mathbb{Q}[\sqrt{2}]$ is $1 = 1 + 0\sqrt{2}$.

Finally, we find the multiplicative inverses of elements in $\mathbb{Q}[\sqrt{2}]$. Let $a, b \in \mathbb{Q}$, with at least one of a or b non-zero, and consider $a + b\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$. We wish to find $c, d \in \mathbb{Q}$ such that $(a + b\sqrt{2})(c + d\sqrt{2}) = 1$.

Observe that $(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2$ is in $\mathbb{Q}$, and is non-zero since $\sqrt{2}$ is irrational. Thus,

$$\frac{(a+b\sqrt{2})(a-b\sqrt{2})}{a^2-2b^2} = 1.$$

If $a^2 - 2b^2$ were 0 with at least one of a and b non-zero then b must be non-zero and so $\frac{a^2}{b^2} = 2$. This gives the contradiction that $\sqrt{2} = \frac{a}{b} \in \mathbb{Q}$.

Therefore, we may take

$$c = \frac{a}{a^2-2b^2}, \quad d = \frac{-b}{a^2-2b^2},$$

noting that this choice satisfies $c, d \in \mathbb{Q}$, and therefore $c + d\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$. Hence, the inverse of $a + b\sqrt{2}$ is

$$\frac{a}{a^2-2b^2} + \frac{b}{2b^2-a^2}\sqrt{2}.$$

Solution to Exercise 2.1

(a) Working in $\mathbb{Q}[x]$:

(i) $f(x) + g(x) = 9x^4 + 2x^3 + 7x^2 - x - 2$

(ii) $f(x)g(x) = 18x^7 + 14x^5 - 29x^4 + 2x^3 - 21x^2 + 3x - 3$.

(b) Working in $\mathbb{Z}_7[x]$, we can reduce the answer to part (a) modulo 7:

(i) $f(x) + g(x) \equiv 2x^4 + 2x^3 + 6x + 5 \pmod 7$

(ii) $f(x)g(x) \equiv 4x^7 + 6x^4 + 2x^3 + 3x + 4 \pmod 7$.

(c) Working in $\mathbb{Z}_2[x]$, we can reduce the answer to part (a) modulo 2:

(i) $f(x) + g(x) \equiv x^4 + x^2 + x \pmod 2$

(ii) $f(x)g(x) \equiv x^4 + x^2 + x + 1 \pmod 2$.

Solution to Exercise 2.2

(a) $\deg(f) = 2$

(b) $\deg(g) = 1$

(c) $$\begin{aligned} f(x)g(x) &= (2x^2 + x + 3)(3x + 1) \\ &= 2 \cdot 3x^3 + (2 + 3)x^2 + (3 \cdot 3 + 1)x + 3 \\ &\equiv 5x^2 + 4x + 3 \pmod 6. \end{aligned}$$

From this, we see that $\deg(fg) = 2$. Note that this is less than $\deg(f) + \deg(g) = 3$.

Solution to Exercise 2.3

(a) $f(x) + g(x) = (2x^3 + 4x^2 - 1) + (3x^3 + 3x^2 - 2x) = 5x^3 + 7x^2 - 2x - 1 \equiv 2x^2 - 2x - 1 \pmod 5$. Therefore $\deg(f + g) = 2$.

(b) $f(x) + g(x) = (1 + x) + (1 - x) = 2$, and therefore $\deg(f + g) = 0$.

Solution to Exercise 2.4

(a) Since $g(x)$ is an associate of $f(x)$, there exists a unit $u \in F[x]$ such that $g(x) = uf(x)$. By Lemma 2.6, $\deg(g) = \deg(uf)$ $= \deg(u) + \deg(f)$, and since $\deg(u) = 0$ by Lemma 2.9, we have $\deg(f) = \deg(g)$.

(b) Let $f(x)$ be a polynomial of degree $\deg(f) = n$ over F with leading coefficient $a_n \neq 0$. Since $a_n \in F$, there exists a multiplicative inverse $a_n^{-1} \in F$ of a_n, and so the polynomial $a_n^{-1}f(x)$ is an associate of $f(x)$, and has leading coefficient equal to 1. That is, the polynomial $a_n^{-1}f(x)$ is monic.

Moreover, it is the unique monic associate: a_n^{-1} is the unique inverse of a_n in F, and so any other associate of $f(x)$ cannot be monic.

Solution to Exercise 3.1

(a)

$$\begin{array}{r|llllll} & x^2 & - & x & + & 2 & \\ \hline x-1 & x^3 & - & 2x^2 & + & 3x & - \ 1 \\ & x^3 & - & x^2 & & & \\ \hline & & - & x^2 & + & 3x & - \ 1 \\ & & - & x^2 & + & x & \\ \hline & & & & & 2x & - \ 1 \\ & & & & & 2x & - \ 2 \\ \hline & & & & & & 1 \end{array}$$

Thus $q(x) = x^2 - x + 2$ and $r(x) = 1$.

(b)

$$\begin{array}{r|llllll} & 2x^2 & & - & 2 & & \\ \hline x^2+1 & 2x^4 & & & & - \ x & + \ 1 \\ & 2x^4 & & + & 2x^2 & & \\ \hline & & & - & 2x^2 & - \ x & + \ 1 \\ & & & - & 2x^2 & & - \ 2 \\ \hline & & & & & - \ x & + \ 3 \end{array}$$

Thus $q(x) = 2x^2 - 2$ and $r(x) = -x + 3$.

(c)

$$\begin{array}{r|llllll} & \frac{3}{2}x^2 & - & \frac{7}{4}x & + & \frac{7}{8} & \\ \hline 2x+1 & 3x^3 & - & 2x^2 & & & + \ 1 \\ & 3x^3 & + & \frac{3}{2}x^2 & & & \\ \hline & & - & \frac{7}{2}x^2 & & & + \ 1 \\ & & - & \frac{7}{2}x^2 & - & \frac{7}{4}x & \\ \hline & & & & & \frac{7}{4}x & + \ 1 \\ & & & & & \frac{7}{4}x & + \ \frac{7}{8} \\ \hline & & & & & & \frac{1}{8} \end{array}$$

Thus $q(x) = \frac{3}{2}x^2 - \frac{7}{4}x + \frac{7}{8}$ and $r(x) = \frac{1}{8}$.

Solution to Exercise 3.2

(a) Since g divides f, there exists a polynomial $a \in F[x]$ such that $f = ag$. Now, using the distributivity of addition and multiplication in $F[x]$, we have

$$f + gh = ag + gh = g(a + h),$$

which shows that g is a factor of $f + gh$.

(b) Since g divides f, we can write $f = ag$ for some $a \in F[x]$. Similarly, since h divides g, there exists $b \in F[x]$ such that $g = bh$. Now, by associativity of multiplication in $F[x]$,

$$f = ag = a(bh) = (ab)h,$$

which shows that f is divisible by h.

(c) Since h divides f and g, we can write $f = ch$ and $g = dh$ for some $c, d \in F[x]$. Now, using associativity of multiplication and the distributive laws, we see that h divides $af + bg$:

$$af + bg = a(ch) + b(dh) = (ac)h + (bd)h = (ac + bd)h.$$

(d) If f and g are associates then there exists $a \in F$ such that $g(x) = af(x)$, which immediately shows that g is divisible by f. Moreover, we can write $a^{-1}g(x) = f(x)$, which shows that f is divisible by g.

For the other direction, suppose g divides f and f divides g. Then there exist polynomials $a, b \in F[x]$ such that $g = af$ and $f = bg$. We will have finished if we can show that a and b are inverses, since this would imply $a \in F$, so that g is an associate of f. We write

$$f = bg = b(af) = (ba)f.$$

Then by Lemma 2.10 we have $1 = 1_x = ba$, and therefore $a = b^{-1}$ as required.

Solution to Exercise 3.3

(a) Since f and g are coprime, there exist $a, b \in F[x]$ such that $1_x = af + bg$. Multiplying this through by h gives

$$h = afh + bgh.$$

Next, as f divides h there exists $c \in F[x]$ such that $h = cf$. Similarly there exists $d \in F[x]$ such that $h = dg$, and we substitute these in to the right-hand side to get

$$h = af(dg) + bg(cf) = (ad + bc)fg,$$

which shows that fg divides h.

(b) Since $\mathrm{hcf}(f, g) = 1_x$, there exist $a, b \in F[x]$ such that $1_x = af + bg$. We multiply this through by h to get

$$h = afh + bgh.$$

Next, since f divides gh, there exists $c \in F[x]$ such that $gh = cf$. Substituting this into the above equation gives

$$h = afh + b(cf) = (ah + bc)f,$$

which shows that f divides h.

Solution to Exercise 3.4

(a) Using polynomial long division, we apply the Euclidean Algorithm:

$$(x^3 + 2x^2 - x - 2) = (x + 6)(x^2 - 4x + 3) + (20x - 20)$$
$$(x^2 - 4x + 3) = \tfrac{1}{20}(x - 3)(20x - 20) + 0.$$

Hence in $\mathbb{Q}[x]$,
$\mathrm{hcf}(x^3 + 2x^2 - x - 2, x^2 - 4x + 3) = \frac{1}{20}(20x - 20) = x - 1$.

(b) From the above calculation,
$x - 1 = \frac{1}{20}\left((x^3 + 2x^2 - x - 2) - (x + 6)(x^2 - 4x + 3)\right)$ and so
$x - 1 = \frac{1}{20}(x^3 + 2x^2 - x - 2) - \frac{1}{20}(x + 6)(x^2 - 4x + 3)$. That is,
$s = \frac{1}{20}$ and $t = -\frac{1}{20}(x + 6)$.

Solution to Exercise 3.5

(a) First, we divide $f(x) = 2x^3 + 3x^2 + 2x - 1$ by $g(x) = x^2 + x$ using polynomial long division:

$$f(x) = (2x + 1)g(x) + (x - 1).$$

Next, we divide $g(x)$ by the remainder, $x - 1$:

$$g(x) = (x + 2)(x - 1) + 2.$$

This tells us that $\mathrm{hcf}(f, g) = \frac{1}{2} \cdot 2 = 1$.

(b) Working backwards through the Euclidean Algorithm,

$$\begin{aligned} 2 &= g(x) - (x + 2)(x - 1) \\ &= g(x) - (x + 2)\left[f(x) - (2x + 1)g(x)\right] \\ &= -(x + 2)f(x) + (2x^2 + 5x + 3)g(x). \end{aligned}$$

Dividing by 2, this gives $a(x) = -\frac{1}{2}x - 1$, and $b(x) = x^2 + \frac{5}{2}x + \frac{3}{2}$.

Solution to Exercise 3.6

(a) First, polynomial long division gives

$$x^4 - x^3 + x^2 - 1 = (x - 1)(x^3 - 1) + (x^2 + x - 2).$$

Dividing $x^3 - 1$ by the remainder gives

$$x^3 - 1 = (x - 1)(x^2 + x - 2) + (3x - 3).$$

The remainder here is $3x - 3 = 3(x - 1)$ and we can see by inspection that $x^2 + x - 2 = (x - 1)(x + 2)$, which allows us to conclude that the hcf is $x - 1$. Therefore, by Proposition 3.12,

$$\begin{aligned} \mathrm{lcm}(x^3 - 1, x^4 - x^3 + x^2 - 1) &= \frac{(x^3 - 1) \cdot (x^4 - x^3 + x^2 - 1)}{x - 1} \\ &= (x^2 + x + 1) \cdot (x^4 - x^3 + x^2 - 1) \\ &= x^6 + x^4 - x - 1. \end{aligned}$$

(b) Dividing one by the other gives

$$x^3 + 5x^2 + 8x + 4 = 1 \cdot (x^3 + x^2 - 8x - 12) + (4x^2 + 16x + 16).$$

The remainder is $4x^2 + 16x + 16 = 4(x^2 + 4x + 4)$, and so we can now divide $x^3 + x^2 - 8x - 12$ by $x^2 + 4x + 4$ to get

$$x^3 + x^2 - 8x - 12 = (x - 3)(x^2 + 4x + 4).$$

Therefore $\text{hcf}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) = x^2 + 4x + 4$.

Finally, using Proposition 3.12,

$$\begin{aligned}\text{lcm}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) &= \frac{(x^3 + x^2 - 8x - 12) \cdot (x^3 + 5x^2 + 8x + 4)}{x^2 + 4x + 4} \\ &= (x - 3) \cdot (x^3 + 5x^2 + 8x + 4) \\ &= x^4 + 2x^3 - 7x^2 - 20x - 12.\end{aligned}$$

Alternatively, note that $x^3 + x^2 - 8x - 12 = (x - 3)(x + 2)^2$ and $x^3 + 5x^2 + 8x + 4 = (x + 1)(x + 2)^2$, and so

$$\text{lcm}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) = (x - 3)(x + 1)(x + 2)^2,$$

which when expanded gives the same result.

Solution to Exercise 4.1

(a) Over $\mathbb{R}[x]$,

$$\begin{aligned}x^4 - 3x^2 + 2 &= (x - 1)(x^3 + x^2 - 2x - 2) \\ &= (x - 1)(x + 1)(x^2 - 2) \\ &= (x + 1)(x - 1)(x + \sqrt{2})(x - \sqrt{2}).\end{aligned}$$

(b) Using the answer to part (a), we see that over $\mathbb{Q}[x]$ we have

$$x^4 - 3x^2 + 2 = (x - 1)(x + 1)(x^2 - 2)$$

and $x^2 - 2$ cannot be factorised further as it has no roots in $\mathbb{Q}$.

For small finite fields such as $\mathbb{Z}_5$, it is often easiest to search exhaustively by computing $f(a)$ for every element a in the field.

(c) By inspection, over $\mathbb{Z}_5$ we see that $x = 4 \equiv -1 \pmod 5$ is a root, which gives

$$x^3 + x^2 + 2x + 2 \equiv (x + 1)(x^2 + 2) \pmod 5.$$

The quadratic $x^2 + 2$ has no roots in $\mathbb{Z}_5$, and so cannot be factorised further.

(d) Over $\mathbb{C}$, we find $f(-1) = 0$, and then using long division (or the answer to part (c)) we obtain

$$x^3 + x^2 + 2x + 2 = (x + 1)(x^2 + 2).$$

Finally, $x^2 + 2 = (x + \sqrt{2}i)(x - \sqrt{2}i)$ where $i = \sqrt{-1}$, and so the complete factorisation is

$$x^3 + x^2 + 2x + 2 = (x + 1)(x + \sqrt{2}i)(x - \sqrt{2}i).$$

(e) We find $f(1) = 0$, which gives $f(x) = (x-1)(x^3 - 5x^2 + 7x - 3)$. For the cubic, we find that $x = 1$ is again a root, which yields $f(x) = (x-1)^2(x^2 - 4x + 3)$. Finally, the quadratic has roots $x = 1$ and $x = 3$, and so

$$f(x) = (x-1)^3(x-3).$$

Solution to Exercise 4.2

(a) Let $p(x)$ be a polynomial of degree 1 in $\mathbb{Q}[x]$. Suppose that $p = ab$ for some elements $a, b \in \mathbb{Q}[x]$. Since $\deg(p) = 1$, the equation $\deg(a) + \deg(b) = \deg(p)$ implies that either $\deg(a) = 0$ or $\deg(b) = 0$. Any element in $\mathbb{Q}[x]$ of degree 0 is a unit. So p has no proper factors and is therefore irreducible.

(b) The theory of quadratic equations tells us that solutions to $f(x) = ax^2 + bx + c = 0$ are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If $b^2 - 4ac < 0$, then $\sqrt{b^2 - 4ac} \notin \mathbb{R}$ and so $f(x) = 0$ has no solutions in $\mathbb{R}$. By the Factor Theorem, this implies that $f(x)$ has no linear factors (and therefore no proper factors) in $\mathbb{R}[x]$, and hence it is irreducible.

If $b^2 - 4ac \geq 0$, then $\sqrt{b^2 - 4ac} \in \mathbb{R}$ and so $f(x)$ has two roots in $\mathbb{R}$. This yields the proper factorisation

$$f(x) = a\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right).$$

Solution to Exercise 4.3

(a) By the Rational Root Test, any root $\frac{p}{q}$ must satisfy $p \mid 1$, $q \mid 4$. The possible options are therefore

$$\tfrac{p}{q} = \pm 1, \pm\tfrac{1}{2}, \pm\tfrac{1}{4}.$$

Trying each in turn, we get

$$\begin{array}{ccc} f(1) = 0 & f(-1) = 6 & f(\frac{1}{2}) = 0 \\ f(-\frac{1}{2}) = \frac{3}{2} & f(\frac{1}{4}) = \frac{39}{64} & f(-\frac{1}{4}) = \frac{75}{64}. \end{array}$$

Therefore the roots of $4x^4 - 2x^3 - 2x^2 - x + 1$ in $\mathbb{Q}$ are $x = 1$ and $x = \frac{1}{2}$.

(b) The possible values that could be rational roots are $\pm 1, \pm 2, \pm 4$. Trying each in turn, we get

$$\begin{array}{ccc} f(1) = 0 & f(-1) = 6 & f(2) = 0 \\ f(-2) = 24 & f(4) = 156 & f(-4) = 300. \end{array}$$

Therefore the roots of $x^4 - x^3 - 2x^2 - 2x + 4$ in $\mathbb{Q}$ are $x = 1$ and $x = 2$.

Solution to Exercise 4.4

If $f(x)$ is reducible, then we can write $f(x) = g(x)h(x)$ in $\mathbb{Q}[x]$ for some $g, h \in \mathbb{Q}[x]$ of degree at least 1. As f is a cubic, one of g, h must have degree 1. By the Factor Theorem, it follows that $f(x)$ must have a root in $\mathbb{Q}$. Employing the Rational Root Test, the only possibilities are $\pm 1, \pm\frac{1}{2}$. Checking each of these in turn verifies that none is a root of $f(x)$, and so f has no factor of degree 1. Hence $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Solution to Exercise 4.5

First, the Rational Root Test tells us that the only possible rational numbers that could be roots of $f(x)$ are ± 1, and $f(1) = f(-1) = 1 \neq 0$. Thus f has no rational roots, and so (by the Factor Theorem) it has no linear factors.

Therefore, we investigate whether $f(x)$ can be factorised into two quadratics over $\mathbb{Z}[x]$. Suppose that

$$f(x) = (ax^2 + bx + c)(rx^2 + sx + t)$$

where $a, b, c, r, s, t \in \mathbb{Z}$. Comparing the coefficient of x^4 gives $ar = 1$, and we can assume that a is positive, so that $a = r = 1$. Next, comparing the coefficient of x^0 gives us $ct = 1$, and so $c = t = \pm 1$. Substituting the values for a, c, r and t, the coefficient of x^3 yields $b = -s$, and the coefficient of x^2 gives $-1 = \pm 1 + bs \pm 1$. Combining these to eliminate s, we find $b^2 = 1 \pm 2$.

Neither $b^2 = -1$ nor $b^2 = 3$ has a solution in the integers, and so $f(x)$ is irreducible in $\mathbb{Z}[x]$. It follows by Gauss's Lemma that $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Solution to Exercise 4.6

Remember that 0 is divisible by *every* non-zero integer.

(a) By Eisenstein's Criterion, the prime p divides every coefficient of $x^n - p$ except that of x^n. Moreover, the constant term, $-p$, is not divisible by p^2. Therefore $x^n - p$ is irreducible.

(b) First, we write $f(x) = \frac{1}{15}\left(9x^5 + 25x^4 - 15x^3 + 50x^2 + 30x - 5\right)$. We can now apply Eisenstein's Criterion with $p = 5$ to conclude that the polynomial in the brackets is irreducible. Therefore, by Lemma 4.6, so is $f(x)$.

Solution to Exercise 4.7

(a) $(x+1)^4 + 1 = x^4 + 4x^3 + 6x^2 + 4x + 2$, and so $(x+1)^4 + 1$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion with $p = 2$. Thus by Lemma 4.17, $x^4 + 1$ is irreducible in $\mathbb{Q}[x]$.

(b) $(x+1)^8 + 1 = x^8 + 8x^7 + 28x^6 + 56x^5 + 70x^4 + 56x^3 + 28x^2 + 8x + 2$, and so $x^8 + 1$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion with $p = 2$.

In fact, it can be proved that $x^{2^n} + 1$ is irreducible in $\mathbb{Q}[x]$ for all $n \geq 0$.

Solution to Exercise 4.8

(a) $x^3 + 7x^2 - 14x + 4$ is irreducible. This can be proved using the Rational Root Test, or reducing modulo 7:

$$x^3 + 7x^2 - 14x + 4 \equiv x^3 + 4 \,(\text{mod } 7).$$

Since any proper factorisation of a cubic must include a linear factor, we now test $x = 0, 1, \ldots, 6$ in turn to see whether they are a solution of $x^3 + 4 \equiv 0 \,(\text{mod } 7)$. None is, and therefore $x^3 + 7x^2 - 14x + 4$ is irreducible in $\mathbb{Z}_7[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Lemma 4.19 and Gauss's Lemma.

(b) This polynomial is not irreducible in $\mathbb{Q}[x]$. The Rational Root Test can be used to prove that $x^4 + 4$ has no roots, and therefore no linear factors. It follows that any proper factor must be quadratic. Without loss of generality, we can assume $x^4 + 4$ to be the product of two monic quadratics, that is, $x^4 + 4 = (x^2 + ax + b)(x^2 + cx + d)$. Comparing coefficients then leads to the factorisation

$$x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2).$$

As there are no linear factors, this factorisation must be complete.

Solution to Exercise 4.9

(a) The irreducible polynomials are

$$\begin{array}{cccc} x, & x+1, & x^2+x+1, & x^3+x+1, \\ x^3+x^2+1, & x^4+x+1, & x^4+x^3+1, & x^4+x^3+x^2+x+1. \end{array}$$

The quadratic and cubic polynomials that are reducible (and therefore not listed above) must have a linear factor, and so either $x = 0$ or $x = 1$ is a root.

Notice that any quartic that is reducible but has no linear factors must be a product of irreducible quadratic polynomials. In $\mathbb{Z}_2[x]$, there is only one irreducible quadratic, namely $x^2 + x + 1$. Thus $(x^2 + x + 1)^2 \equiv x^4 + x^2 + 1 \,(\text{mod } 2)$ is reducible, but all other quartics are irreducible or have a linear factor modulo 2.

(b) (i) $x^4 - 3x^3 + 12x^2 - 2x + 5 \equiv x^4 + x^3 + 1 \,(\text{mod } 2)$ and so is irreducible in $\mathbb{Z}_2[x]$, and hence in $\mathbb{Q}[x]$.

(ii) $3x^4 - x^3 + 71x^2 - 9x + 1 \equiv x^4 + x^3 + x^2 + x + 1 \,(\text{mod } 2)$ and so is irreducible in $\mathbb{Z}_2[x]$, and thus in $\mathbb{Q}[x]$.